DOCUMENT RESUME

ED 041 610 LI 002 069

AUTHOR	Stevens, Mary Elizabeth
TITLE	Automatic Indexing: A State-of-the-Art Report.
INSTITUTION	National Bureau of Standards (DOC), Washington, D.C.
REPORT NO	NBS-Monogr-91
PUB DATE	Feb 70
NOTE	298p.
AVAILABLE FROM	Superintendent of Documents, U.S. Government Printing Office, Washington, D.C. 20402 (GPO C13.44:91, $2.25)
EDRS PRICE	EDRS Price MF-$1.25 HC Not Available from EDRS.
DESCRIPTORS	*Automatic Indexing, *Automation, Citation Indexes, *Classification, *Electronic Data Processing, Evaluation, Indexes (Locaters), *Indexing

ABSTRACT

A survey of automatic indexing systems and experiments has been conducted by the Research Information Center and Advisory Service on Information Processing, Information Technology Division, Institute for Applied Technology, National Bureau of Standards. Consideration is first given to indexes compiled by or with the aid of machines, including citation indexes. Automatic derivative indexing is exemplified by key-word-in-context (KWIC) and other word-in-context techniques. Advantages, disadvantages, and possibilities for modification and improvement are discussed. Experiments in automatic assignment indexing are summarized. Related research efforts in such areas as automatic classification and categorization, computer use of thesauri, statistical association techniques, and linguistic data processing are described. A major question is that of evaluation, particularly in view of evidence of human inter-indexer inconsistency. It is concluded that indexes based on words extracted from text are practical for many purposes today, and that automatic assignment indexing and classification experiments show promise for future progress. (Author)

NATIONAL BUREAU OF STANDARDS

The National Bureau of Standards [1] was established by an act of Congress March 3, 1901. Today, in addition to serving as the Nation's central measurement laboratory, the Bureau is a principal focal point in the Federal Government for assuring maximum application of the physical and engineering sciences to the advancement of technology in industry and commerce. To this end the Bureau conducts research and provides central national services in four broad program areas. These are: (1) basic measurements and standards, (2) materials measurements and standards, (3) technological measurements and standards, and (4) transfer of technology.

The Bureau comprises the Institute for Basic Standards, the Institute for Materials Research, the Institute for Applied Technology, the Center for Radiation Research, the Center for Computer Sciences and Technology, and the Office for Information Programs.

THE INSTITUTE FOR BASIC STANDARDS provides the central basis within the United States of a complete and consistent system of physical measurement; coordinates that system with measurement systems of other nations; and furnishes essential services leading to accurate and uniform physical measurements throughout the Nation's scientific community, industry, and commerce. The Institute consists of an Office of Measurement Services and the following technical divisions:

Applied Mathematics—Electricity—Metrology—Mechanics—Heat—Atomic and Molecular Physics—Radio Physics [2]—Radio Engineering [2]—Time and Frequency [2]—Astrophysics [2]—Cryogenics.[2]

THE INSTITUTE FOR MATERIALS RESEARCH conducts materials research leading to improved methods of measurement standards, and data on the properties of well-characterized materials needed by industry, commerce, educational institutions, and Government; develops, produces, and distributes standard reference materials; relates the physical and chemical properties of materials to their behavior and their interaction with their environments; and provides advisory and research services to other Government agencies. The Institute consists of an Office of Standard Reference Materials and the following divisions:

Analytical Chemistry—Polymers—Metallurgy—Inorganic Materials—Physical Chemistry.

THE INSTITUTE FOR APPLIED TECHNOLOGY provides technical services to promote the use of available technology and to facilitate technological innovation in industry and Government; cooperates with public and private organizations in the development of technological standards, and test methodologies; and provides advisory and research services for Federal, state, and local government agencies. The Institute consists of the following technical divisions and offices:

Engineering Standards—Weights and Measures — Invention and Innovation — Vehicle Systems Research—Product Evaluation—Building Research—Instrument Shops—Measurement Engineering—Electronic Technology—Technical Analysis.

THE CENTER FOR RADIATION RESEARCH engages in research, measurement, and application of radiation to the solution of Bureau mission problems and the problems of other agencies and institutions. The Center consists of the following divisions:

Reactor Radiation—Linac Radiation—Nuclear Radiation—Applied Radiation.

THE CENTER FOR COMPUTER SCIENCES AND TECHNOLOGY conducts research and provides technical services designed to aid Government agencies in the selection, acquisition, and effective use of automatic data processing equipment; and serves as the principal focus for the development of Federal standards for automatic data processing equipment, techniques, and computer languages. The Center consists of the following offices and divisions:

Information Processing Standards—Computer Information — Computer Services — Systems Development—Information Processing Technology.

THE OFFICE FOR INFORMATION PROGRAMS promotes optimum dissemination and accessibility of scientific information generated within NBS and other agencies of the Federal government; promotes the development of the National Standard Reference Data System and a system of information analysis centers dealing with the broader aspects of the National Measurement System, and provides appropriate services to ensure that the NBS staff has optimum accessibility to the scientific information of the world. The Office consists of the following organizational units:

Office of Standard Reference Data—Clearinghouse for Federal Scientific and Technical Information [3]—Office of Technical Information and Publications—Library—Office of Public Information—Office of International Relations.

---

[1] Headquarters and Laboratories at Gaithersburg, Maryland, unless otherwise noted; mailing address Washington, D.C. 20234.

[2] Located at Boulder, Colorado 80302.

[3] Located at 5285 Port Royal Road, Springfield, Virginia 22151.

ED 041610

UNITED STATES DEPARTMENT OF COMMERCE • Maurice H. Stans, *Secretary*
NATIONAL BUREAU OF STANDARDS • Lewis M. Branscomb, *Director*

Automatic Indexing:
A State-of-the-Art Report

Mary Elizabeth Stevens
Center for Computer Sciences and Technology
National Bureau of Standards
Washington, D. C. 20234

U.S. DEPARTMENT OF HEALTH, EDUCATION & WELFARE
OFFICE OF EDUCATION
THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL OFFICE OF EDUCATION POSITION OR POLICY.

"PERMISSION TO REPRODUCE THIS COPYRIGHTED MATERIAL BY MICROFICHE ONLY HAS BEEN GRANTED BY
Govt. Printing Off.
TO ERIC AND ORGANIZATIONS OPERATING UNDER AGREEMENTS WITH THE U.S. OFFICE OF EDUCATION. FURTHER REPRODUCTION OUTSIDE THE ERIC SYSTEM REQUIRES PERMISSION OF THE COPYRIGHT OWNER."

LI 002069

National Bureau of Standards Monograph 91

Issued March 30, 1965

Reissued with Additions and Corrections (See Preface), February 1970

For sale by the Superintendent of Documents, U.S. Government Printing Office, Washington, D. C. 20402 (Order by SD Catalog No. C13.44:91), Price $2.25

Foreword
(1970 Edition)

Widespread interest in the use of computers in automatic indexing created a demand for this publication that led to a recent exhaustion of all stock. While updating and revision would have been desirable, other demands have prescribed reissuance with additional material added as appendices. These are a paper updating the field through September 1966 (Appendix B), and bibliographic citations, pertinent to the subjects in the original text, through August 1969 (Appendix C).

Lewis M. Branscomb
Director

Library of Congress Catalog Card Number: 65-60023

Foreword
(1965 Edition)

The Research Information Center and Advisory Service on Information Processing, (RICASIP), which is jointly supported by the National Science Foundation and the National Bureau of Standards, is engaged in a continuing program to collect information and maintain current awareness about research and development activities in the field of information processing and retrieval. An important responsibility of RICASIP is the preparation of state-of-the-art reviews on topics of current interest in various areas of this broad field.

This report is one of a series intended as contributions toward improved interchange of information among those engaged in research and development in this field. The report considers new uses of machines and automatic data processing procedures for the compilation and generation of indexes to the scientific and technical literature.

A. V. Astin, Director

Contents

Page

Abstract 1

1. Introduction 1

  1.1 Definitions and background 2
  1.2 Scope of this study 10
  1.3 Derivative vs. assignment indexing 13

2. Indexes compiled by machine 14

  2.1 Concordances and complete text processing 15
  2.2 Card catalogs, book catalogs, bibliographies and subject index listings prepared by machine 19
  2.3 Tabledex and other special purpose indexes 25
  2.4 Citation indexes 27
  2.5 Machine conversion from one index set to another 38

3. Indexes generated by machine - automatic derivative indexing 40

  3.1 KWIC indexes 40

    3.1.1 Applications of KWIC indexing techniques 41
    3.1.2 Advantages, disadvantages and operational problems of KWIC indexing 55

  3.2 Modified derivative indexing 68

    3.2.1 Title augmentation 68
    3.2.2 Book indexing by computer 71
    3.2.3 Modified derivative indexing - Baxendale's experiments 73

  3.3 Derivative indexing from automatic abstracting techniques 75

    3.3.1 Auto-condensation and auto-encoding techniques of H. P. Luhn 75
    3.3.2 Frequencies of word n-tuples - Oswald and others 79
    3.3.3 Relative frequency techniques - Edmundson and Wyllys, and others 81
    3.3.4 Significant word distances 83
    3.3.5 Uses of special clues for selection 84
    3.3.6 Recent examples of mixed systems experimentation 86

  3.4 Quality of modified derivative indexing by machine 89

4. Automatic assignment indexing techniques 91

  4.1 Swanson and later work at Thompson Ramo Wooldridge 91
  4.2 Maron's automatic indexing experiments 93
  4.3 Automatic indexing investigations of Borko and Bernick 94
  4.4 Williams' discriminant analysis method 97
  4.5 SADSACT 98
  4.6 Assignment indexing from citation data 99
  4.7 Similarities and distinctions among assignment indexing experiments 100
  4.8 Other assignment indexing proposals 105

Page

5. Automatic classification and categorization 106

5.1 Factor analysis 108
5.2 The theory of clumps 110
5.3 Latent class analysis 113
5.4 Examples of other proposed classificatory techniques 113

6. Other potentially related research 114

6.1 Thesaurus construction, use and up-dating 114
6.2 Statistical association techniques 118

6.2.1 Devices to display associations: EDIAC 119
6.2.2 Statistical association factors - Stiles 119
6.2.3 The association map - Doyle and related work at SDC 122
6.2.4 Work of Giuliano and associates, the ACORN devices 124
6.2.5 Spiegel and others at Mitre Corporation 126

6.3 Clues to index-term selection from automatic syntactic analysis 127
6.4 Probabilistic indexing and natural language text searching 132

6.4.1 Probabilistic indexing - Maron, Kuhns and Ray 133
6.4.2 Natural language text searching - Swanson 134
6.4.3 Full text searching - legal literature 135

6.5 Other examples of related research in linguistic data processing 136
6.6 Machine assistance in translations of subject content indications to special search and retrieval language 140
6.7 Example of a proposed indexing-system utilizing related research techniques 142

7. Problems of evaluation 143

7.1 Core problems 145
7.2 Bases and criteria for evaluation of automatic indexing procedures 149

7.2.1 The Cranfield project 150
7.2.2 O'Connor investigations 151
7.2.3 Questions of comparative costs 153
7.2.4 Summary: potential advantages as bases for evaluation 156

7.3 Findings with respect to inter-indexer and intra-indexer consistency 157
7.4 Special factors and other suggested bases for evaluation 160

8. Operational considerations 164

8.1 Questions of input 164
8.2 Examples of processing considerations 168
8.3 Output considerations 171

9. Conclusion: Appraisal of the state of the art in automatic indexing 173

Page

Acknowledgments 182

Appendix A: List of references cited and selected bibliography 183

Appendix B: Progress and prospects in mechanized indexing 223

Appendix C: Selective bibliography of additional references 237

AUTOMATIC INDEXING

A State-of-the-Art Report

Mary Elizabeth Stevens

A state-of-the-art survey of automatic indexing systems and experiments has been conducted by the Research Information Center and Advisory Service on Information Processing, Information Technology Division, Institute for Applied Technology, National Bureau of Standards. Consideration is first given to indexes compiled by or with the aid of machines, including citation indexes. Automatic derivative indexing is exemplified by key-word-in-context (KWIC) and other word-in-context techniques. Advantages, disadvantages, and possibilities for modification and improvement are discussed. Experiments in automatic assignment indexing are summarized. Related research efforts in such areas as automatic classification and categorization, computer use of thesauri, statistical association techniques, and linguistic data processing are described. A major question is that of evaluation, particularly in view of evidence of human inter-indexer inconsistency. It is concluded that indexes based on words extracted from text are practical for many purposes today, and that automatic assignment indexing and classification experiments show promise for future progress.

1. INTRODUCTION

This report of the Research Information Center and Advisory Service on Information Processing (RICASIP) [1] is one of a series intended as contributions to improved cooperation in the fields of information selection systems development, information retrieval research and mechanized translation. In each of these areas, automatic techniques for linguistic data processing are receiving increased attention. This report covers a state-of-the-art survey of current progress in linguistic data processing as related to the possibilities of automatic mechanized indexing. Insofar as has been practical, the survey of the literature on which this report is based has been made through February 1964.

It has concentrated on the major developments in and related demonstrations of automatic indexing potentialities. Examples are also given of indexes compiled by machine and of potentially related research efforts in such areas as natural language text searching, statistical association techniques used for search and retrieval, and proposed systems for concept processing. There are, undoubtedly, various omissions. Neither the inclusion of reports on various specific experiments and techniques nor the omission of others is intended to reflect an endorsement as such of those that are included or an adverse evaluation of those that are not mentioned.

---

[1] Initiated at the instigation of the National Science Foundation. RICASIP is jointly supported by NSF and NBS.

1.1 Definitions and Background

The noun "index" has as its most general meaning "something used or serving to point out, a sign, token, or indication", (American College Dictionary) or "that which shows, indicates, manifests, or discloses; a token or indication" (Webster's International Dictionary, 2nd Edition, unabridged). More specifically, an index is "a pointer or key which directs the searcher to recorded information".[1] The terms "index" and "indexing" have been used in the fields of library science and documentation with reference to the fact that the selection of information pertinent to a particular problem or interest, from all the previously recorded information available, involves problems of decision-making based on less than the full content or text of each of the records being searched.

Short of complete scanning of all the possibly relevant material, it is necessary to select or "distill" condensed representations or surrogates [2] for each item. These surrogates are intended to direct the searcher to the most probably pertinent items in a collection. The operations known as "indexing" thus involve:

(1) Choosing clues that will serve to identify, for purposes of later retrieval, a particular book, document, or other recorded item, and

(2) Either marking on the item itself or recording as a separate item-surrogate the tags, labels, or codes representing these clues.

The second of these two steps can be purely clerical in nature, but the first has been, to date, primarily the result of human intellectual efforts in subject content analysis.

Well-known inadequacies of human indexing operations include both those stemming from man himself and those which result from the volume and the character of the materials with which he deals. On the human side, there are fundamental questions of perception, comprehension and judgment, as well as those of inter-indexer and even intra-indexer consistency. In addition, the indexer is asked to guess in advance what others will ask for, understand, and find relevant on future search. He is even asked, in effect, to anticipate the language of future inquiries. Thus, a somewhat facetious definition of the noun "index" has a considerable sting of truth: "A system of analyzing information in which the method used to choose categories is carefully hidden from the user. An attempt to outguess the future." [3]

The nature of the material to be indexed, especially in the area of scientific information, raises a number of crucial problems. The still increasing spate of production of technical literature and reports poses not only the problems of sheer volume in terms of manpower requirements and time necessary to produce indexes, but also problems of glut in terms of man-hours necessary for the individual scientist to maintain awareness of what is going on in his field. There are major problems created by newly emerging fields of effort, new interdisciplinary areas of interest, and dynamically evolving terminology. Increasing specialization, on the other hand, brings out additional difficulties in finding what has been done elsewhere that might be applicable to one's own work and in avoiding wasteful duplication of effort, with their own attendant problems of terminology.

---

[1] Crane and Bernier, 1958 [144], p. 513.
(Note: Full citations of references are given in the bibliography by author and by numerical order of the figures in brackets.)

[2] See, for example, R. E. Wyllys, 1962 [651], for discussion of the two-fold purposes of condensed representations: to serve a search-tool function on the one hand and a content-revealing one on the other.

[3] Vanby, 1963 [622], p. 143.

All these problems are aggravated by the increasingly critical urgency which should apply to making all useful information available to those who need it as promptly and as selectively as possible. Recognition of this urgency and of the inadequacies of present solutions has therefore prompted consideration of the feasibility of using machines to assist in the indexing process.

The term "mechanized indexing" signifies the accomplishment of some or all of the indexing operations by mechanized means. The term includes the use of machines to prepare and compile indexes, and to sort, assemble, duplicate and interfile catalog cards carrying index entries. In this report, however, we shall be concerned primarily with the area of automatic indexing, that is, the use of machines to extract or assign index terms without human intervention once programs or procedural rules have been established. This term is chosen in preference to auto-indexing as originally suggested by Luhn (1961 [373]) for the reasons set forth by Bar-Hillel, [1] and to machine indexing [2] due to possible confusion with machine tool operations. Automatic indexing has been used by such workers in the field as Gardin (1963 [209]), Kennedy (1962 [310]), Maron (1961 [395]), Swanson (1962 [584]), and Wyllys (1963 [653]).

For obvious reasons, we also subsume under this term any specifically "clerical" (Fairthorne, 1956 [188], 1956 [189], 1961 [190] and hence machinable operations that can similarly be substituted for human intellectual effort. There is nothing that machines can do which people cannot do except for limitations of time, cost, or availability of appropriate resources. Thus, we shall consider "machine-like indexing by people" (O'Connor, 1961 [447]; Montgomery and Swanson, 1962 [421]) as falling properly within the scope of automatic indexing, especially in the sense of "... deciding in a mechanical way to which category (subject or field of knowledge) a given document belongs ... deciding automatically what a given document is 'about'." [3]

The principle of indexing, that is, of using subject-content clues and item surrogates as substitutes for searches based on perusal of the full contents, has a history of several millenia. In ancient Sumaria and Babylon, clay tablets were sometimes covered with a thin clay envelope or sheath that was inscribed with brief descriptions of the contents of the tablet itself (Carlson, 1963 [101]; Hessel, 1955 [268]; Lalley 1962 [343]; Olney, 1963 [458]; Schullian, 1960 [525]). The first known instance of an index list is apparently that of Callimachus in the third century B.C., which was a guide to the contents of some 130,000 papyrus rolls (Olney, 1963 [458]; Parsons, 1952[469]).

---

[1] Bar-Hillel, 1962 [35], p.417.

[2] Bohnert, 1962[69]; Edmundson, 1959 [176]; and others.

[3] Maron, 1961 [395], p. 404.

Application of the indexing principle by use of clerical procedures that today can be accomplished by machine was suggested a little more than a century ago. A British librarian, Andreas Crestadoro, advocated the permutation of the words in titles in 1856, claiming that thus the subject matter index would follow the author's own definition of the contents of his book. He prepared such "concordances of titles" for several different library collections. [1]

Within a generation, punched card machines had been invented, but they were not to be used for library and documentation purposes for some decades yet. [2] Keppel, writing in 1937 of his vision of the library 21 years in the future, says:

> "When it comes to using the cards, I blush to think for how many years we watched the so-called business machines juggle with payrolls and bank books before it occurred to us that they might be adapted to dealing with library cards with equal dexterity. Indexing has become an entirely new art. The modern index is no longer bound up in the volume, but remains on cards, and the modern version of the Hollerith machine will sort out and photograph anything the dial tells it ... " [3]

By 1945, Bush had prophesied Memex [93], and in the 1950 Windsor lectures Ridenour referred to an RCA development, the so-called "electronic pencil", a proposed reading aid for the blind intended to convert printed characters to a suitable coded form. He went on to suggest:

> "... We shall have to arrange for cataloguing to be done by machine, without human interaction except in terms of setting up once for all the system on which the cataloguing is performed... It is only a step from this device (the electronic pencil) to the electronic catalogue, which will read text for itself, recognize key symbols and phrases with which it has been provided, and construct appropriate catalog entries for the text it reads." [4]

It has only been in the past decade or so, however, that there have been any serious efforts directed to the use of machines for automatic indexing. In the period 1957-1958, Luhn first presented and published several provocative papers dealing with such challenging possibilities as "auto-abstracting", "auto-encoding" and "auto-indexing" (Luhn, 1957 [385]; 1958 [374]; 1959 [371]). Luhn's work on the permutation of significant words in titles, abstracts, and complete text, the Keyword-in-Context or KWIC system, also began about this time.[1] Also in 1958, Baxendale published the results of experiments in automatic indexing involving scanning of topic sentences, syntactical deletion processes and automatic phrase selection (Baxendale, 1958 [41]).

---

[1] See Crestadoro, 1856 [146]; see also Farley, 1963 [192]; Linder, 1960 [362]; Metcalfe, 1957 [416]; and Ohlman, 1960 [451].

[2] See pp.19-22 of this report.

[3] See Keppel, 1939 [316], p. 5.

[4] See Ridenour, 1951 [500], p. 26.

With respect to the KWIC and permuted title techniques, several independent approaches were being developed at about the same time as Luhn's. These concurrent efforts were carried out at the Wright Air Development Center (Netherwood, 1958 [437]), the Rocketdyne Division of North American Aviation (Carlsen, et al, 1958 [99]), and the System Development Corporation (Citron, et al 1958 [120]; Ohlman, 1960 [451]).[2] Netherwood's permuted title index to a bibliography on logical machine design involves manual simulation of a machineable method. Although the results were not published until June 1958, the manuscript was submitted in November 1957.[3] The Rocketdyne permuted-title bibliography, on industrial control, is credited by both Henderson (1962 [263]) and Ohlman (1960 [451]) as the first to be produced on computers, the program

---

[1] In a private communication dated March 13, 1963, Luhn provided the following chronology:

May 1957	Routine 1 Program for word isolation within 60 characters per card, written by H. C. Fallon.
1957-1958	Creation of concordances of various scientific papers in the form of cards, each card showing a keyword centrally located within 60 letters worth of the associated phrase. Experimentation with these cards to arrive at thesauri for special fields of interest or study. Idea of automatic indexing by means of significant or keywords in context conceived by H. P. Luhn.
May 1958	Keyword-in-Context Index for titles only initiated by H. P. Luhn and samples produced with Routine 1 Program.
June 1958	Start punching of titles for Keyword-in-Context Index for literature on Information Retrieval and Machine Translation. (Keypunching done by Miss Olive Ferguson.)
August 1958	Simplified version of Routine 1 written by H. C. Fallon for generating Keywords-in-Context Indexes and delivered to Service Bureau Corporation, New York City.
September 1958	First Fdition of Bibliography and Keyword-in-Context Index on Information Retrieval and Machine Translation published by Service Bureau Corporation.
January 1959	Started writing program for improved version of Keyword-in-Context Index, including derived identification code, written by Jr. J. Havender.
June 1959	Second Edition of Bibliography and Keyword-in-Context Index on Information Retrieval and Machine Translation, published by Service Bureau Corporation, including derived identification codes.

[2] See also National Science Foundation's CR&D Report No. 3, [430], p. 39.

[3] Netherwood, 1958 [437] , p. 155, footnote.

having been written by J. T. Madigan.[1] At any rate, both this program and Luhn's KWIC program at IBM were apparently written relatively early in 1958.

Citron et al (1958 [120]) in presenting results of the SDC work and Ohlman in his chronological bibliography of permutation indexing (1960 [451])cite as at least partial predecessors the "rotated file" principles developed at the Chemical-Biological Coordination Center (1954 [112]; Heumann and Dale, 1957 [270] and 1957 [271]; Wood, 1956 [649]). It should also be noted as a matter of historical background that a system for machine manipulation and compilation of permuted title-and-term-index records has been in productive operation since 1952.[2] This earlier effort was not generally known to other investigators and was apparently first reported in the open literature as late as 1961.

Notwithstanding such other efforts, it is conceded by almost all workers in the fields of automatic abstracting and indexing that the major credit for pioneering interest and impetus should be attributed to Luhn and Baxendale. Specific acknowledgements of their "pioneering work" and "first steps" have been made by many investigators both in this country and abroad--for example, Borko and Bernick,[3] Hines,[4] Mooers,[5] Pevzner and Styazhkin,[6] and Wyllys.[7] In particular, the Russian investigator Purto states: "So far as we know H. P. Luhn was the first investigator to suggest the concept of a set of significant words for the consideration of problems in automatic abstracting."[8]

Much of the early effort 1957-58, whether at IBM or elsewhere, was in fact spurred on by the International Conference on Scientific Information (ICSI) held in Washington, D. C., in November, 1958. The printed text of both the Preprints [478] and the final Proceedings [480, 481] was deliberately prepared, over the typographer's objections, so that a double space followed each period ending a sentence, in order to facilitate machine processing of this text. Thus the printers ".... were faced with ... the necessity to prepare the final volume of the Proceedings from these preprints, and to arrange type composition amenable to computer analysis. The latter is an experiment. With an eye to the distant future, the Program Committee wished to make available the monotype punched tapes from the text for statistical studies with computers. We hope some work of this kind will be demonstrated during the Conference. This has caused some compromises in typography..."[1]

---

[1] Carlsen, et al, "Information Control", 1958 [99], p. 20.

[2] Veilleux, 1962 [624], p. 81: "Consumer demand balanced against availability of manpower and machine time were the factors which led to the establishment of the permutation title word indexing project in 1952."

[3] Borko and Bernick, 1962 [77] , p. 3.

[4] Hines, 1963 [273], p 7.

[5] Mooers, 1963 [424] , p. 4.

[6] Pevzner and Styazhkin, 1961 [472] , p. 3.

[7] Wyllys, 1961 [650] , pp. 6-7.

[8] Purto, 1962 [484] , p. 2.

Several pioneering experiments in automatic indexing were applied to this ICSI material. One of these led to the preparation of a permuted keyword index based on titles, subtitles, section and table headings, figure captions, and selected sentences or phrases taken directly from the text (Citron, et al, 1958 [120]). It was prepared using punched card equipment, and the resulting listings were distributed to the Conference participants in November of 1958. Another set of experiments involved trial of the "auto-abstracting" and "auto-encoding" techniques proposed by Luhn (1958 [379]).[2] A computer program potentially applicable to certain ancillary operations which might be involved in automatic indexing was also demonstrated at the time of the ICSI sessions. (Stevens, 1959 [568]).

Much of the rapidly proliferating work in the field of automatic indexing since that time has been inspired directly or indirectly by the results of these experiments using the ICSI material. For example, Dowell and Marshall, discussing early efforts at the English Electric Company, state: "We first became interested in the possibilities of computer produced indexes through Luhn's work at IBM and the early examples of KWIC indexes which were distributed at the time of the Washington Conference..." (Dowell and Marshall, 1962 [159]).[3]

---

[1] "Preprints of papers of the International Conference on Scientific Information," 1958, [478], Preface. (The monotype tapes are in fact still held in the custody of the Research Information Center and Advisory Service on Information Processing, National Bureau of Standards, but difficulties to be discussed later in this report discourage their use.)

[2] See also his "Automated intelligence systems" 1962 [372], note. 11, p. 100: "Papers for this conference were distributed to participants two months ahead for study. By arrangement with the Columbia University Press the Monotype tapes used in publishing these preprints were made available for experimentation. At the conference exhibit, IBM researchers demonstrated the automatic transcription of these Monotype tapes to magnetic tape via punched cards and thence the automatic creation and printout of abstracts by means of electronic data processing equipment at the Space Systems Center in Washington, D. C. All this was done without any human intervention except for the handling of the input and output records. Also, preprinted Auto-Abstracts of Papers of Area 5 of the Conference were made available to participants at the beginning of the conference."

[3] See also R. A. Kennedy, 1962 [310], p. 181: "While automatic indexing in any interpretative and analytical sense is therefore not yet a practical matter, a simpler mode of machine indexing is coming into wide use ... primarily stimulated by the publication in 1958 and 1959 of reports by Ohlman, Hart and Citron and Luhn."

A somewhat premature attempt was made to establish a subscription service for KWIC indexes for a number of journals, for initial distribution beginning January 1, 1959.[1] Called PILOT (Permutation Indexed Literature Of Technology), the proposed service was advertized as "a revolutionary new *totally cross-referenced* index . . . and it will be produced at the speed of light". Figure 1 is a reproduction of a part of the brochure issued in 1958 by Permutation Indexing, Incorporated, Sol Grossman, President, Los Angeles. While, perhaps unfortunately, the number of subscription orders received was not adequate in terms of the ambitious coverage planned, work on permuted title indexing elsewhere did lead rapidly to the publication of such indexes on a production basis.

As of February 1964, there are more than 40 examples of KWIC and other variations of permuted keyword indexing techniques in productive operation or available to the searcher. KWIC-type techniques have also been extended to special one-time index compilations and other applications, as in "automated content analysis" of verbal protocols of psychiatric interviews and group leadership training sessions (Ford, 1963 [198]; Hart and Bach, 1959 [256]; Jaffe 1962 [294] and 1958 [296]; Stone, et al, 1962 [575]).

The same period during which the ICSI was planned and held (1957-1958) was also marked by the first issue of *Current Research and Development in Scientific Documentation* by the National Science Foundation. In it and in subsequent issues, there were reported other early efforts in machine-compiled indexes, in the construction and use of special thesauri, and in indexing and retrieval experiments based on machine processing of text. Thus, for example, punched card methods for compiling printed indexes and announcement lists were under consideration at Bell Laboratories and at Esso Research and Engineering. Special attention was being given to thesauri as early as July 1957 at both Chemical Abstracts Service and the Cambridge Language Research Unit, and at Ramo Wooldridge, "Research on the problems of fully automatic indexing and retrieval based on raw text input to a general-purpose computer is under way."[2]

Nevertheless, as of the present date, the question of the *possibility* of automatic indexing in the sense of the substitution of machineable procedures for human intellectual efforts normally required to identify, categorize, classify, index, select, and list particular items in a collection of items is still moot. Opinions run the gamut from extreme pessimism, "Mechanization of abstracting and indexing is rejected as impractical for the foreseeable future"[3] to enthusiastic optimism, "The conclusion that automatic indexing and cataloging is superior to human indexing and cataloging is both provocative and remarkable." [4]

Borko and Bernick claim that " ... Raw data, i.e., unedited natural language text, can be processed statistically so as to automatically assign index terms to each document and to classify the document into a subject category; this has been demonstrated." [5] On the other hand, Farradane thinks that any form of mechanized processing in indexing

---

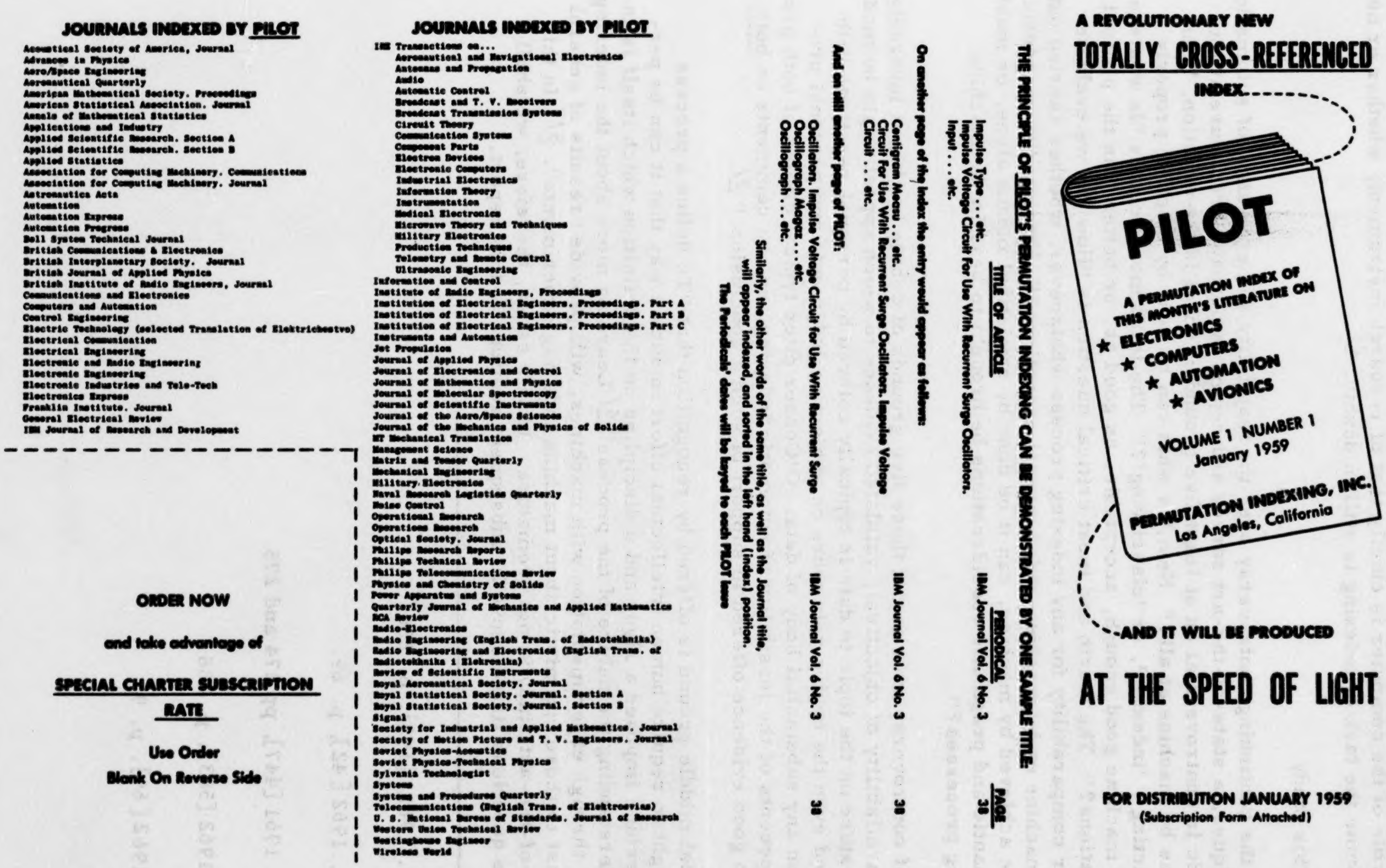
[1] See Linder, 1960 [363], p. 99 and Figure 1.

[2] National Science Foundation's CR&D Reports No. 1, [430] pp. 4, 6; No. 3 [430], pp. 12, 19, 31.

[3] Bar-Hillel, 1958 [33] , abstract.

[4] Swanson, 1962 [584] , p. 468.

[5] Borko and Bernick, 1963 [78] , p. 28.



A REVOLUTIONARY NEW

TOTALLY CROSS-REFERENCED

INDEX

PILOT

A PERMUTATION INDEX OF THIS MONTH'S LITERATURE ON

★ ELECTRONICS

★ COMPUTERS

★ AUTOMATION

★ AVIONICS

VOLUME 1 NUMBER 1

January 1959

PERMUTATION INDEXING, INC.

Los Angeles, California

—AND IT WILL BE PRODUCED

AT THE SPEED OF LIGHT

FOR DISTRIBUTION JANUARY 1959

(Subscription Form Attached)

THE PRINCIPLE OF PILOT'S PERMUTATION INDEXING CAN BE DEMONSTRATED BY ONE SAMPLE TITLE:

TITLE OF ARTICLE	PERIODICAL	PAGE
Impulse Type . . . etc.		
Impulse Voltage Circuit For Use With Recurrent Surge Oscillators.	IBM Journal Vol. 6 No. 3	38
Input . . . etc.		

On another page of the index the entry would appear as follows:

TITLE OF ARTICLE	PERIODICAL	PAGE
Centigram Measu . . . etc.		
Circuit For Use With Recurrent Surge Oscillators. Impulse Voltage	IBM Journal Vol. 6 No. 3	38
Circuit . . . etc.		

And on still another page of PILOT:

TITLE OF ARTICLE	PERIODICAL	PAGE
Oscillators. Impulse Voltage Circuit For Use With Recurrent Surge	IBM Journal Vol. 6 No. 3	38
Oscillograph Magaz . . . etc.		
Oscillograph . . . etc.		

Similarly, the other words of the same title, as well as the Journal title, will appear indexed, and sorted in the left hand (index) position.

The Periodicals' dates will be keyed to each PILOT issue

JOURNALS INDEXED BY PILOT

Acoustical Society of America, Journal
Advances in Physics
Aero/Space Engineering
Aeronautical Quarterly
American Mathematical Society. Proceedings
American Statistical Association. Journal
Annals of Mathematical Statistics
Applications and Industry
Applied Scientific Research. Section A
Applied Scientific Research. Section B
Applied Statistics
Association for Computing Machinery. Communications
Association for Computing Machinery. Journal
Astronautica Acta
Automation
Automation Express
Automation Progress
Bell System Technical Journal
British Communications & Electronics
British Interplanetary Society. Journal
British Journal of Applied Physics
British Institute of Radio Engineers, Journal
Communications and Electronics
Computers and Automation
Control Engineering
Electric Technology (selected Translation of Elektrichestvo)
Electrical Communication
Electrical Engineering
Electronic and Radio Engineering
Electronic Engineering
Electronic Industries and Tele-Tech
Electronics Express
Franklin Institute. Journal
General Electrical Review
IBM Journal of Research and Development

JOURNALS INDEXED BY PILOT

IRE Transactions on...
  Aeronautical and Navigational Electronics
  Antennas and Propagation
  Audio
  Automatic Control
  Broadcast and T. V. Receivers
  Broadcast Transmission Systems
  Circuit Theory
  Communication Systems
  Component Parts
  Electron Devices
  Electronic Computers
  Industrial Electronics
  Information Theory
  Instrumentation
  Medical Electronics
  Microwave Theory and Techniques
  Military Electronics
  Production Techniques
  Telemetry and Remote Control
  Ultrasonic Engineering
Information and Control
Institute of Radio Engineers, Proceedings
Institution of Electrical Engineers. Proceedings. Part A
Institution of Electrical Engineers. Proceedings. Part B
Institution of Electrical Engineers. Proceedings. Part C
Instruments and Automation
Jet Propulsion
Journal of Applied Physics
Journal of Electronics and Control
Journal of Mathematics and Physics
Journal of Molecular Spectroscopy
Journal of Scientific Instruments
Journal of the Aero/Space Sciences
Journal of the Mechanics and Physics of Solids
MT Mechanical Translation
Management Science
Matrix and Tensor Quarterly
Mechanical Engineering
Military Electronics
Naval Research Logistics Quarterly
Noise Control
Operational Research
Operations Research
Optical Society. Journal
Philips Research Reports
Philips Technical Review
Philips Telecommunications Review
Physics and Chemistry of Solids
Power Apparatus and Systems
Quarterly Journal of Mechanics and Applied Mathematics
RCA Review
Radio-Electronics
Radio Engineering (English Trans. of Radiotekhnika)
Radio Engineering and Electronics (English Trans. of Radiotekhnika i Elekronika)
Review of Scientific Instruments
Royal Aeronautical Society. Journal
Royal Statistical Society. Journal. Section A
Royal Statistical Society. Journal. Section B
Signal
Society for Industrial and Applied Mathematics. Journal
Society of Motion Picture and T. V. Engineers. Journal
Soviet Physics-Acoustics
Soviet Physics-Technical Physics
Sylvania Technologist
Systems
Systems and Procedures Quarterly
Telecommunications (English Trans. of Elektrosvias)
U. S. National Bureau of Standards. Journal of Research
Western Union Technical Review
Westinghouse Engineer
Wireless World

ORDER NOW

and take advantage of

SPECIAL CHARTER SUBSCRIPTION

RATE

Use Order

Blank On Reverse Side

Figure 1. Brochure for Proposed Permuted Title Index Service

operations is "liable to continuous error", [1] while Baxendale takes a middle ground: "Thus far the role of the computer is chiefly that of research instrument; whether or not it can fully assume the task of indexing is still in doubt". [2]

1.2 Scope of This Study

In view of the continuing controversy over the feasibility and evaluation of automatic indexing techniques, a state-of-the-art survey and report is perhaps premature at this time. The topic is controversial on at least five grounds: First is the question, "Can indexing be done by machine at all?" Next,"Is what can be done by machine properly termed 'abstracting', 'indexing', or 'classifying'?" The third moot point is "Is whatever can be done by machine good enough, acceptable, as good as, or better than the product of human operations?" The fourth and most critical question is "How can we evaluate acceptability or comparability for any indexing process whatsoever, whether carried out by man or by machine or by machine-aided manual operations?" Finally, "If an indexing product is to be achieved by machine, can it be done by statistical means alone, or must syntactic, semantic and pragmatic considerations be brought to bear in the machine decision-making processes?"

The heat of controversy over any of these five grounds of debate is almost inversely related to the availability of objectively validated evidence to which appeal might be made. Thus, the literature on the topic to date is typically colored by personal reactions both pro and con, and even the cynics rely more on subjective judgments and personal preferences than on any substantial body of data. O'Connor cites typical claims of both proponents and opponents of the feasibility of automatic indexing, and he comments on both, "I have seen no good evidence offered in support of such a conclusion." [3]

An impartial middle ground is offered by recognition that "To define a process ordinarily thought to require human intellectual effort in such a way that it can be performed by a machine imposes a rigor and a discipline on the definition which itself is invaluable to understanding the nature of the process".[4] Learning more about the indexing process itself, through experimentation with machines, will provide "results of general interest, not just to those optimistic about machine indexing experiments". [5] In this sense, a state-of-the-art study is not premature. In this sense, therefore, we shall explore the five questions listed above in subsequent sections of this report.

---

[1] Farradane, 1961, [193], p. 236.

[2] Baxendale, 1962 [42], p. 69.

[3] O'Connor, 1961 [447], pp.274 and 275.

[4] Swanson, 1962 [583], p. 288.

[5] Bohnert, 1962 [69], p. 9.

More particularly, in this survey of automatic indexing efforts, we will be concerned with the following principal topics:

(1) A brief indication of the variety of ways in which punched card machines and computers can be and have been used in the preparation or compilation of indexes. [1]

(2) A more detailed consideration of the possibilities for machine generation of indexes, specifically including:

  (a) Automatic derivative indexing, as in various examples of machine extraction of keywords, where selection is based upon pre-specified criteria,

  (b) Automatic assignment indexing, whereby the machine is programmed to determine, in accordance with various specified criteria, whether or not some one or more members of an established list of 'labels' (such as subject headings, class names, descriptors, or other indexing terms) should appropriately be assigned to the document or item in question, and

  (c) Automatic classification techniques, on which such assignment-indexing operations may or may not be based.

(3) Consideration of the use of machines as relatively sophisticated aids to human intellectual operations applied in either subject-content analyses or search-strategy determinations.

(4) Discussion of the question of evaluation of any index whatever, whether manually or mechanically prepared.

(5) Consideration of the implications of related research and development efforts, specifically including:

  (a) Comparative evaluation of indexing systems,

  (b) Development and use of new types of "indexing" aids (in the sense of "pointing to" and "indicative of" the probable subject-content relevance) to either selective dissemination or retrospective search of the technical literature,

  (c) Linguistic and logical-inference approaches to the elucidation of 'meaning' in natural-language messages, and

  (d) Theoretical approaches to the problems of determining "membership-in-classes".

---

[1] Note that card-controlled camera systems, such as the Listomatic, and Addressograph machines have also been used for index compilations. See, for example, Shaw, 1951 [542], p. 49, who cites early use of the Addressograph for bibliographical work by A. Predeek, "Die Adrema-Maschine als Organizationsmittel im Bibliotheksbetriebe", Berlin, 1930, and E. Morel, "Les Machines au secours de la Bibliographie", Revue du Livre 1:14-19 (1933). Use of such devices is not included in this report, however, since they cannot be adapted to machine generation of indexes.

(6) Appraisal of the current prospects for further research and development.

Certain difficulties of organization are evident. Thus many proposals precede actual tests of techniques to which they are akin. Other proposals have been engendered as by-products of or incidental to investigations of other techniques, such as those of text processing to derive by machine selected sentences which together may serve as automatically generated "abstracts", more properly extracts. [1]

This related subject of automatic abstracting, i. e., the application of machine-usable rules to the extraction or generation of textual information representing in condensed form that carried in the document as a whole, will not be of primary concern. However, it will be noted that most of the automatic abstracting techniques so far proposed are potentially usable as tools for automatic indexing, especially in the trivial sense that the automatic selection of index terms could be based solely upon the substantive words found in the machine-prepared extract. [2] Further, since we are presuming that a state-of-the-art review of automatic indexing techniques is in some sense appropriate at this time, we shall emphasize the actual results of machine compilation and machine generation of indexes and those investigations of assignment-indexing techniques for which experimental or comparative data have been reported, rather than theoretical approaches.

---

[1] See, for example, Luhn, 1959 [384], p. 4: "The principle of abstracting information by extracting certain portions or elements from the full text of a document is particularly suitable to mechanization"; Becker, 1960 [44], p. 13: "Perhaps 'extracting' would have been a better word than 'abstracting'"; Edmundson and Wyllys, 1961, [181], p. 227: "All proposed methods for making an automatic abstract of a document involve using the author's own words by selecting complete sentences, thereby reducing abstraction to the simple task of extraction."

[2] See Wyllys, 1963 [653], p. 22: "Automatic indexing is an area that seems to us to be especially close to automatic abstracting, since the words and word groups found to be most representative of a document for automatic abstracting purposes are obvious candidates for entries in an automatic index for the documents." See also Tanimoto, 1961 [594], p. 235: "Thus after extracting k sentences which are a predetermined small fraction of the document, we have an 'abstract'. To find the indexes to the document we take these k sentences and the corresponding sets of the canonical elements and consider terms versus sentences instead of sentences versus terms... The same analysis is then applied to this 'transposed' problem to produce the index terms"; Yakushin, 1963 [654], p. 17: "If some method can be employed for the automatic compilation of abstracts, it can as well be used for the subject index."

1.3 Derivative vs. Assignment Indexing

At least part of the provocation and controversy with respect to the possibilities for the use of machines in indexing is due to confusion as to what type of indexing is meant. This in turn relates to a much older and broader controversy--that between "word" or "catchword" indexing on the one hand and "subject indexing", "concept indexing", or "controlled indexing" on the other.

In terms of operational definition, the contrast is best expressed in Luhn's distinction between index entries that are derived from the text of an item itself and those that are assigned to it from a list or schedule of subject categories, descriptors and the like, which exists independently of the text of the item (Luhn, 1962 [372]). 1/ In general, the differentiations that are made for the broader controversy, and the claims and counter-claims made by the enthusiasts of either school, provide background for the distinctions that should be made between various automatic derivative indexing operations and whatever possibilities may be demonstrated for assignment indexing by machine.

In his text on information storage and retrieval Kent (1962 [315]) contrasts word indexing as used in permuted keyword indexes, concordances and "pure" uniterm systems with controlled indexing which "implies a careful selection of terminology used in indexes in order to avoid, as far as possible, the scattering of related subjects under different headings." He notes elsewhere that word indexing requires little subject-matter training on the part of the indexer and little skill in indexing as such, and adds: "It is this type of indexing that a machine can perform well."2/

Like Kent, Bernier thinks that true subject or assignment indexing requires highly trained human indexers. He says further:

> "The difference between subject and word indexing has been unclear at times. Both types employ words, but only true subject indexing employs them with discrimination. Word indexing leads to omission of entries, scattering of related information, and a flood of unnecessary entries. Word indexing uses words as they are found in the material indexed with a minimum regard for standardized meaning..." 3/

Herner provides a further amplification of differences that are pertinent to consideration of indexing by machine, as follows: 4/

1/ See also Herner, 1962 [266], p. 5; Skaggs and Spangler, 1963, [557], p. 60; Slamecka, 1963 [558], p. 224. Mooers makes a similar distinction between "index terms which are words or phrases extracted from the text and stylized conceptual terms--cliches --which are assigned to the text", 1963 [423], p. 4.

2/ Kent, 1962 [314], p. 268.

3/ Bernier, 1956 [54], p. 23.

4/ Herner, 1963 [267], p. 183.

"The differentiation that is made between the two types of indexing is that word indexing is inextricably tied to the words in a text: If a word appears it gets indexed as such; if it does not appear it does not get indexed. Concept indexing, on the other hand, has an element of abstraction in it: Words may either be indexed as such or may be converted, either by themselves or in combination with other words, into concepts which may not bear a direct resemblance to the words or combinations of words that evoked them in the indexer's mind."

Machine techniques such as those of Luhn's KWIC, like the early Uniterm systems, look no farther than the words used by the one author himself. Techniques such as those of Maron, Swanson, Borko, Meadow and Williams, among others, look specifically to relationships between words as used by one author to patterns of word usages in a given subject area or given document collection. They may also look to these patterns as in turn related to prior human analytic judgments of the "aboutness" referrents of items in the collection. In this sense, they at least attempt replication by machine of assignment indexing.

There is no real question but that machines can in fact derive words from text provided that it is in machine-readable form. This machine procedure may involve direct extraction of all words as index entries, as in a complete concordance. It may involve the extraction of only those words which survive a "purging" operation in which articles, conjunctions, adjectives, and other "common" words are first deleted. Various machine-controlled modifications to such "derivative" indexing are also available. The case for machine achievement of assignment indexing for any but limited special cases is not so clear.

2. INDEXES COMPILED BY MACHINE

A first and obvious use of machines in indexing processes is in the manipulation of index entries, previously selected on the basis of human analysis, to produce various orderings, duplications and listings of these entries. The power of machine techniques to speed and economize the sorting, ordering and listing operations in the preparation or compilation of indexes was recognized quite early, both in the field of library science and in the consideration of potential areas of application by specialists in machine potentialities.

In particular, two specialized types of index, at least in the broad sense, are such that their compilation would be almost prohibitive in terms of time and cost were it not for the use of machines. These are, respectively, the case of the complete index, the index to all words of a text in their various contexts, which is a concordance, [1] and the case of the "citation index", which has been used in the field of law for many years but has only quite recently been suggested for literature search purposes related to scientific and technical information.

---

[1] See, for example, Doyle,1963 [162], p. 11: "Without data-processing machinery, concordances are prohibitively expensive to generate for most uses except in those cases where it is well known that a given volume of text is going to be used again and again, by large numbers of people over a long period of time. As we know, clergymen have made use of manually prepared concordances of the Bible since the 12th century".

In machine-compiled indexes, no item or entries are eliminated by the machine, whereas in even the most rudimentary of machine-generated indexes, such as KWIC, various reductive or extractive operations are automatically applied as a part of the machine procedure. We shall be concerned in this section with brief discussions of machine-compiled indexes and related devices, specifically, concordances, card or book catalogs mechanically prepared, citation indexes, and special indexes such as Tabledex. The use of machines to compile, sort, duplicate and list index entries can only be considered to be mechanized indexing in a relatively trivial sense. We shall consider, therefore, only a few representative examples, emphasizing early work and some of the pioneering instances.

2.1 Concordances and Complete Text Processing

When as early as 1856, Crestadoro proposed the use of permutations of the words in titles as a subject-content index the only "machines" available for the processing operations were people acting in a strictly clerical way. Precisely such clerical operations have been used for centuries in a process that is, in the special sense of full representation of document contents, an index-producing operation--the making of concordances.[1] The task of listing each separate word in a book in all the contexts in which it appears is incredibly time-consuming and tedious when carried out by manual means. There are those who have spent the major part of their lifetimes at this task. For example: "It took James Strong thirty years to compile his exhaustive Concordance of the Bible..." [2] The use of machines capable of processing signals which represent and preserve information offered a potentially revolutionary change, and with the advent of the electronic computer even more radical possibilities of very high speed processing were opened up.

As early as 1949, J. W. Mauchly (the co-inventor of ENIAC and UNIVAC) envisioned the use of computers for documentation and library science activities. He suggested that the full information contents of the Library of Congress collections could be recorded in machine language, stored in this form on magnetic tape, and searched by machine in a procedure which would match words or other selection indicia occurring in the recorded information to the specified words or selection criteria of a query or search prescription. Specifically, he estimated that the entire collection, then amounting to 10,000,000 books, could when transcribed to binary-code representation [3] be serially searched in 20 hours.[4]

---

[1] See, for example, Black, 1962 [65], p. 314: "The oldest book in the world has had such an index for many years--the concordance to the Bible;" Markus, 1962 [394], p. 19: "The ultimate in permutation for indexing is a published concordance;" Linder, 1960 [363], p. 99: "We know of a concordance prepared in the 13th Century;" Simmons and McConlogue, 1962 [555], p. 3: "Complete indexing has been used of course for centuries in the preparation of concordances."

[2] Carlson, 1963 [101], p. 211.

[3] That is, markings which have one of two values (thus, binary digits or "bits"), can be used to distinguish between 2^n different other symbols such as alphabetic characters by using log 2^n of such markings. A binary code for the 26 letters of the English alphabet requires a five-bit representation for each letter. If numeric digit characters are also recorded, (26+10), a six-bit code representation is required.

[4] Mauchly, 1949 [406], p. 295. See also "Report to the Secretary of Commerce on the application of machines..." 1954 [620], p. 67.

Mauchly's suggestion was, in effect, the idea of a complete index that could be searched by machine. We should note, however, that although subsequent technological advances could significantly decrease his original time estimate, the crucial questions that remain are those of what, assuming one-to-one representation of document text, one would search for. [1/] Natural language searching by machine, in the sense of full text inspection, is a "pay-as-you-go" concordance technique. It is, however, a technique which must be aided and abetted by various forms of synonym reduction, syntactic normalization, homograph resolution and other special processing operations if it is to be in any sense an effective tool for selection of clues to be retrieved.

Gardin, in a series of recent lectures on automatic documentation, (Gardin, 1963 [207, 208])[2/] refers to the opinions of some investigators that it should be possible to "jump" the stage of indexing and to search the natural language texts directly. The problem, he points out, then shifts to the determination of all the various ways in which the possible answers to a question may have been expressed in these natural language "complete indexes". Instead of carrying out reductions or condensations of the documents, as in normal indexing procedures, amplifications of questions are required. "Reductive" indexing of the source documents can only be eliminated at the expense of "expansive" indexing of questions. Gardin concludes that the gain from this is very doubtful.

There is also the presently staggering burden of time and cost to convert full texts to machine-usable form. As of February, 1961, it was estimated that the natural language text material available for machine processing amounted to little more than the words contained in the Harvard Classics five-foot shelf (Stevens, 1962 [567]). Perhaps up to ten times that amount is now available, notably in the 6,000,000 words of the statutes of Pennsylvania [3/] and in several million additional words that have since been keypunched at the Center for Automation of Literature Analysis, Gallarate, Italy. [4/] A very recently

---

[1/] See, for example, Yngve, 1959 [657], pp.978-979: "We will have to find formal connections between widely divergent ways of saying essentially the same thing. In addition there is much that we will have to learn about searching. If we had today a complete grammar of English which was capable of rendering explicit all the relations and distinctions implicit in the document, I doubt that we would know how to use it effectively in a machine search situation. We would be embarrassed by the very wealth of the information available. Much more must be learned about search situations."

[2/] See also Bar-Hillel, 1962 [35], p.415: "Could not the stage of clue assignment be completely skipped and the request topic be directly compared with the original documents? It is very natural that such a thought should have arisen, but it must be stressed that there is nothing in our knowledge of the workings of communication which would indicate that such a proposal is, or ever will be, practical."

[3/] See various references by J.F. Horty, W.B. Eldridge and S.F. Dennis, E.M. Fels, R. Wilson.

[4/] R. Busa, data reported at the NATO Advanced Study Institute on Automatic Document Analysis, Venice, July 1963.

completed study made by the TRW Computer Division, Thompson Ramo Wooldridge, involves the investigation of the possibilities for a center to provide text in machine-usable form. The report gives a total figure of approximately 50,000,000 words of text so available as of February 28, 1964, but this includes non-scientific text, such as newspaper and popular magazine materials (Mersel and Smith, 1964 [415]).

Mersel and Smith also report on the estimated requirements for machine-usable text for various research groups, averaging over a million words per year per group. Yet, at present keypunching costs of one cent or more per word, is it reasonable to assume that any of these research groups can provide a budget of over $100,000 per year for this purpose alone? Moreover, this budget would provide for the conversion of no more than a thousand 1,000-word items or a hundred 10,000-word items at costs, respectively, of $100 or $1,000 per item. For the present, therefore, the conclusion is inescapable: either indexing or search based upon full text processing is not yet practical. Even the most enthusiastic proponents of "searching full natural language text" (Swanson, 1960 [589]) and "maximum-depth indexing"(Simmons and McConlogue, 1962 [555]) generally agree as to the present impracticality of full-text mechanized indexing except for special limited cases.

The two problems of determining what to search for, given full text, and of feasibility of conversion of text into machine-usable form thus combine to limit "complete indexing" largely to the special cases of providing corpora for studies in the field of computational linguistics and of compiling the traditional scholarly tool--the concordance to all the words in a given literary work or works. Apparent exceptions, including experimental work with abstracts only and the law statutes studies, are usually cases in which the selective principle of disregarding common words (and hence the bulk of the actual text) is applied automatically either on input or in subsequent processing (Cleverdon and Mills, 1963 [131]). These cases, therefore, may be considered machine-generated indexes rather than machine-compiled. Moreover, it should be noted that:

> ".....The law, itself, is an appropriate field for data retrieval. The statutes, especially, are written in relatively clear, concise language. At least, this is their intent. Practically, this means that input and output can both be relatively short and that retrieval of legal information will be involved with fewer semantic difficulties." 1/

In the area of concordance-making, however, the potentialities of machine compilation have been put to good use. The pioneer efforts in this area are unquestionably those of Father Roberto Busa, S.J., of the Gallarate Center. As early as 1946, Busa proposed to his superiors that a card file recording all the words used in all of the works of St. Thomas Aquinas should be set up, and he began his actual experiments using IBM punched card equipment in 1949 (Busa, 1953 [87], 1960 [91], and 1958 [92]; Secrest, 1958 [540]). 2/ Appearing in 1951, his Sancti Thomas Aquinatis Hymnorum Ritualium Varia Specimina Concordantiarum is the first known example of a complete word index that was compiled by machine techniques. The early Gallarate work was carried out on standard punched card equipment, but from the time of the concordance to the Dead Sea Scrolls, computers have also been used (Tasman, 1959 [595], [596], and [597]). The major continuing task is still to other works of St. Thomas. Other machine-compiled concordances produced by Busa's Center include one to Goethe's Farbenlehre, Bd. 3.

1/ Asher and Kurfeerst, 1963 [24], pp.1-2.

2/ See also Scheele (ed.), 1961 [522], pp.206-209.

Other relatively well-known examples of machine-compiled concordances include those to the Revised Standard Version of the Bible (Ellison, 1957 [186]; Cook, 1957 [139]) and to Matthew Arnold's poetry (Painter, 1960 [461]; Parrish [467, 468]). The Cornell Concordance Series, under the general editorial supervision of Parrish, includes investigations of Old English, such as The Anglo-Saxon Poetic Records (Bessinger, 1961 [59]).

The November 1962 issue of Current Research and Development in Scientific Documentation, No. 11, [430], lists several concordances compiled by machine including the work of Sebeok [533, 534] and associates at Indiana University on Cheremis folksongs, the work on the National Vocabulary of the French language under Quemada at the University of Besancon, [1] the preparation of glossaries and concordances to the works of Kant at the University of Bonn [2] , and concordances to medieval German texts being compiled by Wisbey at the University of Cambridge (Wisbey, 1962 [646], [647]). At the University of Gothenburg in Sweden, work has begun on mechanical linguistic analysis of English language texts, using the machine-readable teletypesetter tapes used for the printing of paperback books (Ellegård, 1960 [184] and 1962 [185]). [3] Another recent example is that of the work at the Summer School of Linguistics, University of Mexico (Grimes and Alvarez, 1961 [243]). By 1963, Marthaler writes that "Compiling concordances with the aid of a computer is already standard routine to such an extent that it needs hardly be described in detail." [4] As of January 1964, a general-purpose computer program for the IBM 7090 which can compile various types of concordances has been announced as available from the Mechanolinguistics Project at the University of California. (1964 [95]). [5]

The major advantage of using machines to compile concordances is, of course, the enormous difference in the time required to complete the work. Thus, only 120 hours were required on the UNIVAC computer to prepare the 800,000 words of the Concordance to the Revised Standard Version of the Bible (Cook, 1957 [139]; Ellison, 1957 [186]). [6]

---

[1] See "Actes du colloque sur le mecanisation...", 1961 [1]; Quemada, 1961 [485] and 1959 [486]; Centre d'Etude du Vocabulaire Francaise, "Specimens de Travaux lexicographiques...", 1960 [106].

[2] National Science Foundations CR&D Report No. 11 [430] p. 316

[3] Ibid, p. 321.

[4] Marthaler, 1963 [399] , p. 14

[5] "California Concordance Program Available", 1964 [95]

[6] Carlson, 1963 [101] , p. 211.

In the use of the IBM 705 for the concordance to the Summa Theologiae, Fr. Busa reports that only 60 hours were required to arrange in alphabetical order 1,600,000 words. 1/ This advantage of speed, with the concomitant benefits of both economy and timeliness, is illustrated by Tasman as follows:

> "... It has been estimated that it would take 50 scholars 40 years... to manually index the 13 million or so words of St. Thomas Aquinas' complete works. IBM punched card machines would produce the indexes and concordances much more accurately and would take ten scholars about four years. Large-scale data processing techniques would reduce the time to about 25 percent...(or)...ten scholars to do the job in less than a year." 2/

Other advantages stem from the facility with which further machine processing can be introduced. Once the text is in machine-readable form, a number of valuable byproducts can be derived. Examples are statistics on the number of words that have 2, 3, ... n letters, frequencies of letter usage; printouts of occurrences of specified words or groups of words; and lists alphabetized on terminal rather than initial letters. Added advantages of computer processing are further exemplified in the options available with the California concordance computer program (1964 [95]), some of which are as follows:

(1) The user may obtain a restricted rather than a full concordance by supplying a list of words for which no entries are to be made.

(2) The user may obtain a selective concordance by supplying a list of words for which, and only for which, entries are to be made.

(3) Each entry word may be centered with its preceding and succeeding context, up to the limits of one full line of 131 characters, or each entry word may be listed together with the full sentence or verse in which it occurs.

(4) Text with interlinear information such as grammatical symbols can be used and selective concordances can be compiled on the basis of such interlinear information.

(5) The citations of an entry can be listed in order of textual occurrence, in an order determined by preceding or following words in its context or in an order determined by accompanying interlinear symbols.

2.2 Card Catalogs, Book Catalogs, Bibliographies and Subject Index Listings Prepared by Machine

The use of machines such as punched card equipment for the preparation and processing of library card catalogs and of index listings was advocated by a few far-sighted documentalists at least as early as the 1930's (Parker, 1938 [463]; Dewey, 1959 [153]).

1/ See his statement in Scheele, 1961 [522], p. 209.

2/ Tasman, 1958, [596] , p. 11.

McCormick's bibliography on mechanized library processes (1963 [407]) lists a number of early suggestions, notably those of Fair in 1936 [187], Shera in 1938 [547], and Gates [225] and Callander [96, 97, 98] in 1946. Cox, Bailey and Casey proposed the use of punched card equipment for the preparation of bibliographies in the field of chemistry in 1945 [142].

By 1946, Gull claimed that:

> "...Punched cards and present equipment offer new possibilities right now for solving the problems of the indexes to Chemical Abstracts. These indexes are large undertakings in themselves, and the work of arranging, cumulating, and printing them can be simplified by placing the index information on punched cards at the time the abstracts are made. With current indexes on punched cards, two or three cumulations of the author index during the year will greatly reduce the work required in using current issues from that approach. Cumulations of the subject, patent, and formula indexes immediately become possible for intervals more frequent than once a year." [245]

The following year (1947) saw a summary by Gull of potential applications of punched cards in special libraries [247], and Becker surveyed some of the then discernible prospects for library mechanization, as a student in the Library School of Catholic University. He stressed such advantages as flexibility in the processing of new material for abstracting, indexing, filing, and interfiling purposes and the printing out of various listings in any format. [1]

The potential use of machines for library science and documentation had not actually been recognized, however, for many years after the invention of punched card equipment. Both the punched card developments (beginning with Hollerith and Powers in the 1880's) and the electronic computers developed from 1946 onward were first applied to the automatic manipulation of information in the sense of statistical, mathematical, or engineering data, rather than to information about data or information about other information. Dr. John Shaw Billings, himself a librarian of note, was apparently the first to suggest to Herman Hollerith the idea of recording information as holes punched in cards which could then be sorted mechanically. [2] Larkey comments: "It is not known if Billings ever thought of applying the principle to bibliographic work, but it would seem eminently fitting that it might be so utilized." [3]

Larkey himself as head of the Army Medical Library Research Project at the Welch Medical Library, Johns Hopkins University, was certainly one of the pioneers in such utilization, but this was almost 70 years from the date of the Billings-Hollerith conversations. The Army Project, begun in late 1948 or early 1949, had as its contract objective "to explore existing and projected methods, emphasizing machine methods, applicable to such pilot projects as may be necessary" (Larkey, 1949 [348], 1956 [349], and 1953 [351]). Also as of 1949, the Library of the Department of Agriculture is reported to have "conducted an experiment in the use of electronic data-processing machines to produce the author and subject indexes to the 'Bibliography of Agriculture'." 1/

---

[1] Becker, 1947, [43], pp. 11-12: "From the flexible arrangement of the cards, bibliographies become readily available by subject, author, and title. In special libraries, where material on one subject is concentrated, the research possibilities of gathering, sorting, filing, and printing information are almost limitless. Continuous machine interfiling permits keeping current with new entry additions."

[2] "With the masters...", 1963 [648], p. 18.

[3] Larkey, 1953 [351], p. 34.

It is not until the early 1950's, however, that punched card machine techniques were actively put to use for the preparation of card catalogs, book catalogs, bibliographies and various index listings. Then, a number of independent but largely concurrent applications were tried out on at least an experimental basis, including in addition to the work of the Welch Medical Library Project pioneering efforts in mechanized book catalog production (Griffin, 1960 [242]; Martin, 1953 [400]; Berry, 1958 [58]) and what is claimed to be the "first successful non-experimental punched-card catalog of periodicals", the Serial Titles Newly Received (now New Serial Titles), as published by the Library of Congress from 1951 onwards. 2/

The work at the Welch Medical Library continued for several years, the final report being issued in 1955 [234]. Beginning in 1951, the project maintained in punched card form the subject heading authority list used for the Current List of Medical Literature (Larkey, 1953 [351]; Garfield, 1953 [217] and 1954 [220]." Garfield has stated that this work "clearly demonstrated the ease of converting alphabetic subject heading lists to categorized or classified lists of terms by the use of punched card equipment." 3/ That is, each heading or subheading had assigned to it a numeric code reflecting its appropriate position in the classified system, which could then be used by machine for sorting, ordering and listing. Ingenious use was made of the IBM 101 Statistical Machine in the preparation of printed subject indexes (Garfield, 1953 [218] and 1954 [216]). Other subject heading lists maintained by punched card techniques by 1953 or earlier included those of the U.S. Patent Office and the Technical Information Division of the Library of Congress. 4/

The first loose-leaf printed book catalog to be produced by machine methods was apparently that of the King County Public Library in the State of Washington in 1951, and the following year the Los Angeles County Library inaugurated a similar system for the distribution of a master book catalog prepared by mechanized techniques (Berry, 1958 [58]; Griffin, 1960 [242]; Martin, 1953 [400]; Alvord, 1952 [4]).

The work on mechanized preparation of lists of periodicals at the Library of Congress has been reported as follows:

> "In 1951, the Library began publishing, at monthly intervals, Serial Titles Newly Received. In 1953, its title was changed to New Serial Titles... Ever since its inception, the fundamental ingredient of the publication has been the IBM punched card...

1/ U.S. Congress, Senate Committee on Government Operations, 1960 [619], p. 147.

2/ Dewey, 1959 [153], p. 36.

3/ Garfield, 1959 [221], p. 471.

4/ Garfield, 1954 [220], p. 1.

> "Two important advantages of the punched-card method were foreseen when the publication began. First, it would be possible to print lists from the cards at will, without any further editing or proofreading, once the information was in punched-card form. Second, there was the possibility of mechanically preparing special lists of titles, selected on the basis of subject, country, or language." [1]

Thus, by 1953, "a number of instances of printed indexes prepared by machine" could be claimed. [2] The use of punched cards to sort, to prepare tabular listings for various drafts and revisions, and to interfile corrected or revised entries greatly facilitated the preparation at Battelle Memorial Institute of the subject index to the Proceedings of the International Conference on the Peaceful Uses of Atomic Energy, 1955 (Lipetz, 1960 [367]).

Developments in the use of punched card machine techniques in bibliographic operations of these types, beginning in the 1950's, have by no means been limited to the United States. For example, Remington Rand punched cards have been used in the preparation of a national union catalog of Italian libraries, [3] and Mikhailov reports for the All-Union Institute of Scientific and Technical Information (VINITI) as follows:

> "The development program for machine production of indexes has been underway at the Institute for a number of years...In fact, operational use of Soviet-made punch-card machines to compile the author indexes for some of the series of our Abstract Journal has been practiced at the Institute since 1957." [4]

In France, at the Centre d'Etudes Nucleaires, Saclay, a program has been developed for mechanization of the production of biweekly and cumulative indexes and for demand searches (Chonez, 1960 [116, 117, 118]).

With the advent of automatic data processing systems, the speed, the flexibility and the capability for multiple-purpose processing buttress the claim that the card catalog can be "replaced or supplemented by book catalogs made with the aid of mechanized equipment". [5] It is further claimed that "The printed catalog produced by means of automatic equipment combines the best features of the conventional card catalog and the traditional printed catalog, and adds to both new dimensions that would have been unbelievable a generation ago." [6] A joint project is under way by the Medical Libraries of Columbia,

---

[1] U. S. Congress Senate Committee on Government Operations, 1960 [619], p. 85.

[2] Larkey, 1953, [351], p. 38.

[3] Berry, 1958 [58], p. 287.

[4] Mikhailov, 1962 [410], p. 50.

[5] McCormick, 1963 [408], p. 195.

[6] Vertanes, 1961 [625], p. 242. This is with reference to the LILCO Library Printed Catalog, which is prepared by sorting and processing information on titles, authors and titles-by-subject-groupings serving as indexes to the holdings at the Long Island Lighting Company.

Harvard, and Yale Universities for computer preparation of book catalogs for books published from 1960 onward (Kilgour, et al 1963 [324]). Another recent illustrative example of the production of printed book catalogs by means of computer compilation is that of the Boeing "SLIP" System (Weinstein and Spry, 1963 [633]).

Along with recognition of computer-processing potentialities there has emerged increased awareness of the desirability of taking advantage of one-time recording of information to serve multiple purposes: the principle of by-product data generation. The advantages for the library and document collection are that a single recording of bibliographic information in machine-usable form can lead to a variety of products, specifically including printed book catalogs, 1/ recurrent and demand bibliographies, the requisite number of copies for conventional card catalogs, card catalog sets or catalog listings for the personal use of the individual worker, input to mechanized selection and retrieval systems, and machine-manipulatable data for such other purposes as circulation control.

Turner and Kennedy report, for example, the initial use of a Flexowriter to prepare library catalog cards and the by-product generation, via a 1401 computer, of bi-weekly listings of unclassified report titles at the Lawrence Radiation Laboratory, the "SAPIR" System (Turner and Kennedy, 1961 [615]). Chasen discusses a change from a previous punched card system for circulation and recall at General Electric's Missile and Space Division Laboratory to a combined Flexowriter and G. E. 225 computer procedure to provide mechanized retrieval, compilation of desk catalogs, computer updating of catalogs and files, and the maintenance of subscription lists (Chasen, 1963 [108]).

Fasana describes a system at the Air Force Cambridge Research Laboratory Library where typing indications in the tape are used as boundary codes. He reports:

> "Input tapes are currently being processed on a computer to automatically produce catalog card sets, circulation control records, and book form indexes. Original input tapes now being accumulated will form the basis of a machine-searchable file to be used in the future for more sophisticated printouts and searches." 2/

For such applications, Durkin and White make the following typical claims:

> "The system described has permitted the IBM Command Control Center Engineering Library to produce its catalog cards and library bulletin both faster and cheaper. Since a by-product of this process is the preparation of all catalog information in

---

1/ See for example, Olney, 1963 [458], p. 42: "During the past few years a number of libraries have initiated a program of mechanization... by punching on IBM cards or paper tape some of the bibliographic information normally given on catalog cards. Recording this information in machine-readable form makes it very easy to prepare printed book catalogs..."

2/ Fasana, 1963 [195], p. 326. This system involves the "Machine-Interpretable Natural Format" and procedures developed for AFCRL by Itek Corporation; see also Lipetz et al, 1962 [368].

punched card form, it has also permitted the establishment of a circulation control system, the publication of overdue notices and reading lists, and the eventual institution of a computer retrieval program" (Durkin and White, 1961 [173]; White, 1963 [638]).

Heiliger reports for the library of the new Chicago Campus of the University of Illinois as follows:

"The type of bibliography the computer can produce does make greater use of LC card information than do present card catalogs. With the computer programmed with a set of library filing rules and a set of symbols that describes for the computer the various parts of the bibliographic unit, it can print-out, for instance, a list of books published in a given country, between certain years, on a certain subject (or combination of subjects), that are illustrated and have bibliographies. It will also be possible to permute on individual items in LC subject headings in the same fashion that Chemical Titles does on titles. This index has been dubbed POSH (permuted on subject headings)." [1]

Some recent experimental work at Inforonics, Inc. puts major emphasis on by-product data generation, beginning with the actual preparation of manuscripts for publication. Tape typewriter processing of manuscript for journal articles is being studied from the point of view of producing machine-usable text. This text, together with coded identification of the separate items in the text, is so prepared that computer programs can produce from the single-input automatic typesetting tapes for the article itself, author and subject index entries, and the like. Computer text transformations can also produce entries for citation indexes, abstract journals and search files (Buckland, 1963 [83, 84]).

Other computer-produced indexes or special indexes involving compilation rather than selection by machine include indexes to Nuclear Science Abstracts (Day and Lebow, 1960 [151]), the Current List of Medical Literature (Chonez, 1960 [116, 117, 118]), the Retrieval Guide to Thermophysical Properties Research Literature, [2] and the Research and Development Abstracts of the USAEC (Sherrod, 1963 [541]). At the Atomic Energy Commission also, a modification of this RDA computer program is used for author, corporate author, number and subject indexes for the Engineering Materials List, which includes announcements of blueprints and drawings. [3] In several instances, machine processing capabilities are used for permuted listings under various assigned indexing terms. [4] Special cases of machine permutation operations involve compilation and organization of chain indexes, used to reflect the various key entries in faceted classification systems (Dowell and Marshall, 1962 [159]; Foskett, 1962 [199]; Olney 1963 [458]).

---

[1] Heiliger, 1962 [259], p. 475.

[2] Markus, 1962 [394], p. 19; Touloukian, 1962, 1963 [607].

[3] Davis, 1963 [150] p. 237.

[4] See, for example, reports on the SWIFT program for NASA's STAR (Newbaker and Savage, 1963 [438]); the AIMS System (Heller, 1963 [260], and the SPINSTRE System (Wheater, 1963 [639]).

A final special case of a computer-compiled index should be noted. This is the work of Schultz and Sherpherd with reference to the annual meetings of the Federation of American Societies for Experimental Biology (FASEB) (Schultz and Shepherd, 1960 [532]; Schultz, 1963 [527]; Shepherd 1963 [545]". [1] The indexing terms are generated first by the authors of the papers but are then run against a computer program, which by thesaurus-type look-up eliminates synonyms and supplies syndetic devices in addition to formatting the subject index for printout.

The machine-readable thesaurus developed for this project presently performs the following four basic functions (Schultz, 1963 [527]):

1. It accepts words from titles and indicia supplied by the authors without modification if they match acceptable indexing terms.
2. It recognizes certain other words as acceptable if modified and modifies them accordingly, for example, by "use" directions for synonyms and near-synonyms.
3. It adds additional indexing terms when certain words occur, an example being "'penicillin', use also 'antibiotics'."
4. It deletes certain words if they do not occur in the context of an acceptable indexing phrase.

2.3 Tabledex and Other Special Purpose Indexes

The uses of machine techniques in index compilation so far discussed represent instances in which conventional tools of bibliographic control can be prepared at lower cost or more rapidly, or both. In addition, however, certain new and unconventional types of index have been or are being produced with the aid of computers.

The Tabledex method, as proposed by Ledley in 1958 (Ledley, 1958 [352], Zusman, et al, 1962 [661]; O'Connor, 1960 [442]), involves coordinate indexing in bound book form, with special features to facilitate search, conserve space and display index terms co-occurring with a given term for a given item. [2] A major advantage claimed for this method is that by the use of computers bibliographies and book-form indexes can be organized, compiled, and printed in page format within a matter of hours.

A Tabledex index typically consists of a bibliography proper, in which each citation has been assigned an identifying number; an alphabetical list of the indexing terms used, which may also have numeric codes; and a set of indexing tables. These tables contain item numbers in the leftmost column, and either the names or the codes for indexing terms assigned to an item along the row. There is one such table for each distinct term used in indexing the items.

---

[1] These investigators claim the first production of a conventional subject index by computer.

[2] See, for example, O'Connor, 1960 [446], p. 241: "Ledley approximately halves the average size of the document descriptions required by imposing an order on the vocabulary of indexing terms. When a document description belongs in a term subset, only those terms of the description need to be recorded which come later in term order than the term of the term of the subset. This illustrates another type of storage organization."

To facilitate searching, only those terms which are of higher numeric or alphabetic order than that for the term for which the particular table is compiled are recorded in the rows. Thus to make a search on several terms, the user turns to the table for the one of these terms that has the lowest term value, which table records all items to which the term has been assigned, and checks the rows of the table for the second lowest ranking term, the third, and so on. Variations in the Tabledex method allow for the automatic assignment of numeric codes to the indexing terms based on relative frequency of use within the collection. Ledley also discusses methods for finding articles associated with all except one, all except two, or all except n of the given words in a search prescription.[1]

A first example of a computer-compiled Tabledex index was that to a bibliography prepared by the Library of Congress for the International Geophysical Year (Zusman et al, 1962 [661]).[2] The computer program for the IBM 7090 carried out the operations of assigning accession numbers, extracting index terms and compiling the term lists, determining frequencies so as to assign frequency numbers to the terms, organizing and preparing the tables, and developing an author index. Two formats were used, one giving terms by numeric code and the other spelling out the terms as normal words. The latter feature provides a measure of browsability in the system.[3] A Tabledex compilation program is also in use at the Applied Physics Laboratory of Johns Hopkins University (Olmer and Rich, 1963 [454]).

Another coordinate index search tool, making use of what is in effect a document-descriptor matrix with special codes and column arrangements to save space and facilitate rapid scanning, is the Scan-Column Index suggested in 1960 by O'Connor [449]. He further suggested the use of computers for compilation, as follows:

> "A computer can organize information about documents into a scan-column index. The input needed consists of the document identifications and their accompanying

---

[1] Ledley, 1959 [352], pp. 1235-1239.

[2] See also National Science Foundation CR&D No. 11 [430], pp. 130-131.

[3] Zusman, et al 1962, [661], p. ii: "... The word tables have the advantage that browsing can be accomplished and possible associations made during the search... Such 'browsing' can be enhanced by including at the end of each row in a table all the other words also associated with the article of that row".

index terms. . . and an indication of either the number of columns desired or the column density desired. The computer will determine the frequency of each term, the positive and negative correlations of terms, and the quantity of these correlations by counting or sampling key figures, such as the average number of terms per document. It then can assign column-character codes accordingly."[1]

In 1961, Costello described the use of computer techniques for compilation and computer printout of a dual dictionary for a coordinate indexing system using links and roles at DuPont's Polychemicals Department. After manual analysis, term-role assignments are keypunched, the cards are listed for editing including the elimination of synonyms and the indication of appropriate postings to more generic terms, and re-keypunched for conversion to magnetic tape. Tapes for posting of items and links to term-roles are merged by computer with tapes giving alphabetical equivalents of term codes and with appropriate syndetic indications for final output on an IBM 407 high-speed printer [141].

Still another instance of a coordinate index, modified to show pre-coordination of terms as compiled by computer, is that at the Electronic Properties Information Center (Johnson, 1963 [301]). The system consists of abstract cards maintained in accession number order, together with machine printouts that pre-coordinate descriptors within nine major categories. The listings of pre-coordinated descriptors are arranged in three different indexes; alphabetically arranged within each category, alphabetized without respect to category but with code indication of the category reference, and a non-categorized listing arranged alphabetically in reverse order. Advantages of machine processing include the ease with which various statistical counts can be made, such as the average number of items in the system for a given material and a specified property. Summary indications of the state-of-the-art in the field of interest can be obtained, "for the system will indicate not only areas where research has been done, but also areas where gaps in the literature occur, and a measure of the growth of research activities in the field can be developed."[2]

2.4 Citation Indexes

"A citation index is a directory of cited references in which each reference is accompanied by a list of source documents which cite it."[3] This is a relatively new type of bibliographic search tool that would be almost impossible to compile without the use of machines.[1] In at least one case, moreover, the availability of mechanical devices was itself the inspiration for the idea of a citation index to the scientific literature. Garfield states in a 1954 paper that he was led to the idea of "Shepardizing" from an earlier concern with the development of citation codes or "coden"[2] that would facilitate machine processing of bibliographic and index entries.[3]

---

[1] O'Connor, 1962 [449], pp 18-49.

[2] Johnson, 1963 [301], p. 296.

[3] Sher and Garfield, 1963 [546], p. 63.

The value of *Shepard's Citations* in tracking down precedents and decisions has been recognized in the legal field for many years.[4] The desirability of a similar tool for literature searchers in the fields of scientific and technical information was suggested about a decade and a half ago, when Seidell and others proposed its use for patent searching (Seidell, 1949 [541]; Hart, 1949 [255]). In 1954, the Bush Committee in its considerations of the potential applicability of machines to Patent Office problems received a proposal from the Atlantic Research Corporation of Alexandria, Virginia, which was to cover "the development of a Patent Citation Index, comparable to Shepard's Citations".[5] In the period 1954-1956, both Garfield[6] and Fano[7] independently advocated the development of a citation indexing tool for scientific and technical literature. As of today, there are at least five or six instances of citation indexes that have been produced, several different experimental investigations are under way, and new interest has been generated by the considerations of the Weinberg Panel. Thus:

---

[1] See, for example, Atherton, 1962 [25], p. 4: "The volume of data to be processed is so massive that processing machines are a necessity"; Garfield 1954 [210], p. 4: "Where such large volume of data is to be handled it must be expected that mechanical devices of high speed and versatility... would probably be a determining factor in the system's success."

[2] That is, brief codes, often mnemonic, for journal title abbreviations and other clues to publisher and date of publication.

[3] Garfield, 1954 [210], p. 2.

[4] *How to Use Shepard's Citations* [281] has been published periodically by Shepard's Citations, Inc., Colorado Springs, since 1873.

[5] U.S. Dept. of Commerce "Report to the Secretary of Commerce...," 1954 [620], p. 27.

[6] Garfield [210, 211, 212]. Adair, writing in January, 1955, specifically acknowledges a suggestion of Garfield's (for 1955 [2], p. 32) but Garfield in turn credits Adair, (1963 [214], p. 290).

[7] Fano, 1956 [191], p. 3: "Let us accept, at least for the sake of this argument, the conclusion that linguistic associations between documents cannot lead to a satisfactory definition of a bibliography. Then the only other type of association for which evidence is available is that provided by simultaneous references in the literature, by the concomitant use of documents by experts as evidenced by library records, and by other similar joint events."

> "Of the newer approaches to the indexing of scientific documents, the Weinberg Panel was particularly impressed with the citation index as a promising bibliography tool. In order to learn more about this approach, the National Science Foundation is currently sponsoring the compilation and publication of extensive citation indexes for the fields of genetics and also for statistics and probability; and is supporting two kinds of experiments to evaluate different techniques for using citation data in indexes and searching systems in the field of physics." [1]

In general, the principle of citation indexing is based upon the hypothesis that the bibliographic references cited by an author provide significant clues to the subject content of the author's own paper and/or that there is a certain commonality in subject between papers that cite the same references or that are co-cited. [2] The principle can be applied to the compilation of bibliographical or indexing tools in several different ways. First, there is the method of citedness, which groups for a given item the identifications of subsequent items that have cited it. The converse of this is, of course, the bibliography or reference list of a given item. [3] In the first case, we are concerned with "descendants," and in the list of references with "ancestors". [4]

---

[1] Committee on Scientific Information, 1963, [135], p. 16.

[2] Compare Adair, 1955, [2], p. 32, with respect to Shepard's Citations itself: "Since all of the cases listed under a given case have cited it, it follows that they must all be, more or less, pertinent to the case cited." See also Kessler, 1963, [320], p. 1: "This method ... originated in the hypothesis that the bibliography of technical papers is one way by which the author can indicate the intellectual environment within which he operates, and if two papers show similar bibliographies there is an implied relation between them."

[3] See Salton, 1962, [520], p. III-3: "A citation index consists of a set of bibliographic references (the set of 'cited' documents), each being followed by a list of all those documents (the 'citing' documents) which include the given cited document as a reference. A citation index is to be distinguished from a reference index which lists all cited documents under each citing document."

[4] See, for example, Tukey, 1962, [611], p. 5: "Any user's greatest need is likely to be for access to the latest information rather than to the oldest, but the latest items are children, not ancestors. Genealogy is important, but progress requires tracing descendants Iung and Vandeputte, 1960, [291], p. 11, make a similar distinction between "histoire" (antecedents) and "filiation" (successors).

A second method, implied in Fano's suggestions for the use of relative frequencies of association between items found in the literature, is one of citingness, which groups together items that cite one or more identical references. This method has been developed by Kessler and his associates as the technique of "bibliographic coupling" (Kessler, [317] through [323]. The purpose here is to identify groupings of related items where relatedness is defined in terms of the number of references shared by each of the members of the group with some given test paper or with each other. It is noted that where the citedness index and the reference list typically give the bibliographic references themselves as the searching or retrieval tool, the bibliographic coupling technique seeks rather to define groups of similar papers. 1/ A third method, and one which may be combined with either of the other two, is to derive indexing terms for a given paper from the overlay of indexing terms previously assigned to any papers which it cites. Salton 2/ further suggests that:

> "... Citation indexes could be used to extend a given set of index terms by starting with the terms attached to a given document or document set, and adding to them the 'related' terms obtained from new documents which cite the original ones."

The suggested advantages of citation indexing include the claims that this tool does not require trained indexers, 3/ that it is highly susceptible to mechanization (Garfield, 1955 [213], 1956 [212], 1957 [211]; Atherton, 1962 [25]; Becker and Hayes, 1963 [45]), and that it may cost significantly less than subject indexing. 4/ A major advantage claimed is responsiveness to user, rather than indexer, interests and view points. 5/ Some of the representative claims with respect to this factor are as follows:

1/ See Atherton and Yovich, 1962 [26], p. 3: "Kessler's method, however, does not retrieve the references cited by a paper. Instead these references are examined to determine the 'bonds' between papers; e.g., if two papers share six references, in common, they are said to have a 'coupling strength' of six. By applying either of two criteria of coupling, one can 'filter out smaller groups of papers' related to a given paper."

2/ Salton, 1962 [520], p. III-8; see also Lesk, 1963 [356].

3/ Atherton, 1962, [25], p. 3.

4/ See Atherton and Yovich, 1962 [26], pp. 3-4: "Garfield estimates cost of abstracting and indexing 200,000 articles in one year to be $3 million. He estimates the cost of a citation index for these same articles (approximately 3 million citations) to be $300,000." See also Doyle, 1963, [162], p. 8: "The editing labor, the input preparation cost, and the automatic processing time are all so small that it's very likely citation indexing is destined for a great surge of popularity in the immediate future."

5/ Committee on Scientific Information, 1963 [135], pp. 55-56: "Because the indexing is based on the author's rather than on an indexer's estimate of what articles are related to what other articles, citation indexes are particularly responsive to the user's, rather than to the indexer's viewpoint."

"The most feasible scheme for alerting individuals to what is of interest in their own field requires an on-going up-to-date citation index. For each narrow field of interest of an individual there are, it is believed with good reason, three to five to ten key items such that:

(c1) If he knew that a new item referred to one of his key items, the individual would be glad to skim the new item,

(c2) An individual who skimmed all new items referring to one of his key items would be adequately alerted to the newest results in his own specialties." [1]

"A research worker who finds one article several years old can relate later developments by locating all subsequent articles that have referred to it. Corrections and errata can be brought together by a citation index." [2]

"Citation indexing will overcome artificial dividing lines that are drawn in various abstracting services." [3]

"It is believed that citation indexes will be useful...in bringing together related materials in different fields where the interrelationships are not readily identifiable from other types of indexes." [4]

"Since the end product of a citation indexing is a listing which collects in one place the bibliographical descendants of a given cited author, bringing these titles together helps to illuminate for the searcher the extent and nature of information association patterns employed by other authors who had a similar or related interest to his own. Its development, therefore, serves as an approach to the user's frame of reference, not the indexer's." [5]

The importance of being able to pick up more than the principal subject matter clues is indeed an advantage of citation indexing. Garfield, commenting on the potential cross-breeding of interests, gives an example of a personal search for more information on the RCA electronic scanning pencil in which he was led to one of Busa's reports on machine use in philological analysis and to an article of interest in the field of information theory. [6] Garfield further points out that the cross-breeding can extend across changes of terminology with time,[1] and Lipetz suggests that it can break down barriers with respect to use of foreign literature.[2]

---

[1] Tukey, 1962 [611], p.9.

[2] Atherton, 1962 [25], p.2. See also Garfield, 1955 [213], p.1.

[3] Atherton and Yovich, 1962 [26], p. 3.

[4] Brownson, 1963 [82], p. 3. See also Garfield, 1957 [211], p.4.

[5] Becker and Hayes, 1963 [45], p.137.

[6] Garfield, 1954 [210], pp.4-5.

Other claimed advantages relate to the usefulness of the citation index for purposes other than those of direct literature search. Such other purposes include identification of significant research by "equating frequency of citation with relative significance of subject matter", (Salton, 1962 [520]), determinations of the number of references cited in a given field or by journal or publication date (Atherton, 1962 [25]), evaluation of the relative importance of various scientific journals (Westbrook, 1960 [636]; Kessler, 1961 [322]), tracing of trends in the history of ideas or in a particular field of literature (Brownson, 1963 [82]; Salton, 1962 [520])[3] and empirical studies of the frequencies of self-citation, multiple authorship, and the like (Atherton, 1962 [25]).

A number of disadvantages of the citation index are to be noted, however. First is the obvious lack of consistency between authors in terms of whether or not they cite the prior literature at all and in terms of the completeness and correctness of the citations they do make.[4] Atherton quotes Westbrook as saying:

> "Science is subject to changing fashions of interest that lead to a distorted number of published papers in a given subject and an inordinately high level of citations to any one who reports first on the fashionable subject. The method will not appraise work performed but not published."[5]

---

[1] Ibid, p. 6: "Changes in terminology are to a certain extent overcome through the citation approach, since the author who makes a reference to a paper that is forty or fifty years old is making the jump in terminology for us." See also Garfield, 1956 [212], p.11.

[2] Lipetz, 1963, [366], p.265: "It is reasoned that availability of a citation index derived from Soviet physics journals and approachable through familar American references should stimulate utilization of the Soviet physics journals in the United States."

[3] See also Reisner, 1963 [497], p. 71: "Citation indexes are receiving increasing attention as bibliographic aids and as sociometric tools. As sociometric tools, they are being used to explore the flow of information across national boundaries and from pure to applied fields, to determine the structure of a field, and to determine the 'value' of documents or authors."

[4] See, for example, Doyle, 1963 [162], p. 8: "The disadvantages of this kind of indexing is, of course, that it depends on authors providing ample and suitable references"; Salton, 1962 [520], p.III-7: "In many cases personal preferences are evident both as to number and types of papers cited; authors have varying backgrounds, and there may also exist a tendency toward self-citation regardless of relevancy"; Thompson, 1963 [600], p. II-1: "The difficulties... are largely due to the extreme variability of format and to the lack of standardization which prevails in the publication of citations."

[5] Atherton, 1962 [25], p. 4, citing J.H. Westbrook.

An author not cited frequently enough or not cited within a given time period will not appear in the citation index. Doyle points out that there are "many kinds of documents we would like to retrieve where it is not customary to provide citations at all". 1/ In the bibliographic coupling method, both those papers which make no references to any other paper and those papers which do not share at least one reference with some other paper in the system are automatically excluded. 2/

Other disadvantages of the citation indexing technique relate to difficulties of the lack of standard practices in the citing of references and to problems of recognizing whether one citation is or is not equivalent to another. These are, of course, related to the normal difficulties arising from non-standardized formats and practices in descriptive cataloging, in use of journal abbreviations, in transliterations of foreign language titles and names, and the like, but they are now aggravated by the present prospects for direct machine processing. As Lipetz points out:

> "Author's names may be cited in somewhat different ways, and there is no simple mechanical procedure for bringing together the different versions. For example, an author's name may be cited both with and without initials; it would take a comparison of the additional information on the cited reference to establish that these authors are the same. Even more difficult are the problems of mechanically determining that a misspelling has occurred." 3/

Both the disadvantages of incomplete and disproportionate coverage and of failures to equate equivalent citations are quite readily obvious to the user of a citation index if he is reasonably familiar with the subject field or document set that is covered. Thus, the use of the citation index as the exclusive tool for literature search is subject to defects of both oversight and 'over-cite' which are cumulative and which are often easily recognizable. Atherton and Yovich emphasize that: "Knowledge of these weaknesses tends to prevent anyone from trusting the system's ability to retrieve the pertinent literature." 4/

In general, however, the citation index has not been proposed as an exclusive means for literature search and retrieval, but rather as one of a set of tools or as a supplement to other indexes. 5/ In this connection, it is of interest to note that a manual technique of literature search tested at The Thermophysical Properties Research Center

---

1/ Doyle, 1963 [162], p. 8.

2/ See Atherton and Yovich, 1962 [26], p. 39; Marthaler, 1963 [399], p. 23.

3/ Lipetz, 1962 [364], p. 262.

4/ Atherton and Yovich, 1962 [26], p. 39.

5/ See, for example, Tukey 1962 [611], p. 10: "The citation index, in its retrieval and pursuit uses, is not something to be used alone. Rather, it is the tool whose presence makes all the other tools more effective."

while not using a citation index as such, makes use of a supplementary citation tracing technique both to shorten manual search time through abstract journals and to follow up additional search leads (Lykoudis, et al, 1959 [387]; Cezairliyan, 1962 [107]). The technique is briefly described as follows:

> "One starts searching the abstracting journal beginning with the most recent issue and going back through a number of years, a. Next, the bibliographies of the papers located in these a years are searched for new references. The references found in this second step of the search will, in general, cover a period of years (b - a). Then one reverts back to searching through the abstracting journal again for another period of a years starting with the year b. This cyclic procedure of alternate searches through the abstracting journal, followed by searching the bibliographies of uncovered papers, is repeated until the total number of desired years of search is covered." 1/

In a sample search on the thermophysical properties of metals, the results showed that the cost of the cyclic procedure was only 65% of the cost of conventional manual search using the abstract journals only.

Recent efforts in the development and use of citation indexes proper include experiments in evaluation at the American Institute of Physics, 2/ an extensive compilation and processing program at the Institute for Scientific Information, 3/ and a cooperative program between the Statistical Techniques Research Group of Princeton University and the Bell Telephone Laboratories (Tukey, 1962 [611] and [612]). Reisner has reported work on the compilation of a citation index to 30,000 patent disclosures and its experimental evaluation in progress at IBM's Thomas J. Watson Research Center (1963 [497]). Goodman is concerned with a citation index to the literature of new educational media, especially that on programmed learning and teaching machines (1963 [235]).

At the Centre d'Etudes Nucleaires de Saclay, a citation index to papers in the field of thermonuclear fusion and plasma physics is being prepared. 4/ Lipetz is carrying on work in the preparation and evaluation of citation indexes, begun at the Itek Corporation, as an independent worker and consultant to the A.I.P. project. 5/ Carroll and Summit report that citation indexing is under consideration at Lockheed's Missile and Space Division, (1962 [102]). Kessler and associates at M.I.T. 6/ and Salton's group at

---

1/ Lykoudis et al, 1959 [387], abstract, p. 351.

2/ Atherton and Yovich, 1962 [26]; National Science Foundation's CR&D Report No. 11, p. 12.

3/ Ibid, pp. 27-28.

4/ Ibid, p. 76.

5/ Ibid, p. 181.

6/ Ibid, p. 128.

the Harvard Computation Laboratory (Salton, 1961 [512], 1962 [513], 1963 [514] and [515]), are concerned with citations as a basis for grouping and categorizing sets of related documents.

Early examples of citation indexes that have been produced include the precedents in the fields of statistics and information theory listed by Tukey. [1] Tukey also refers to early experimentation involving manually manipulated card files by J. L. Hodges, Jr., Charles H. Kraft, and William H. Kruskal. [2] Goodman (1963 [235]) describes the use of Termatrex cards showing for each item other items cited by it.

Examples of machine-compiled citation indexes, however, are those of Garfield and Sher in the field of genetics (1963 [546]), Lipetz's experimental index to the citations in the proceedings of the two United Nations conferences on the peaceful uses of atomic energy, (1961 [364], 1960 [365]), and the citation index to references listed in the "Short Papers" submitted for the 1963 Annual Meeting of the American Documentation Institute (Luhn, 1963 [377]). As of January, 1964, the first five volumes of Science Citation Index are available from the Institute for Scientific Information. These volumes are reported to have 2,250,000 lines of copy representing the computer-compiled citation trails for 102,000 articles published in 1961. [3]

Preliminary evaluations of the citation indexing principle have, as noted previously, been carried out in an American Institute of Physics project supported by the National Science Foundation. One experiment involved the selection of a single paper from the December 1, 1961 issue of The Physical Review and the tracing of references and citations through that journal for the period 1956 to 1960. A bibliography of 64 papers was produced as a result. This was then evaluated by a nuclear physicist, who found that the titles alone were an insufficient basis for judging whether or not these papers should all have been included, and who commented critically that there was no way of knowing if all the papers really relevant to the subject of the test paper had indeed been found. A further check by search of the subject index did in fact reveal six pertinent papers which had been missed by the citation indexing technique.

A second experiment at the American Institute of Physics involved application of Kessler's "coupling strength" criteria to 41 of the 64 papers selected in the first experiment, the remainder being excluded because they shared no references with any other paper. The resultant groupings of presumably highly related papers were also evaluated by a subject matter specialist, who found them relevant to each other but the selection incomplete. Atherton and Yovich, reporting these A.I.P. experiments, concluded that: "More work will have to be done before the usefulness of citation indexing can be accurately determined." [4]

---

[1] Tukey, 1962 [611], pp. 23-24.

[2] Ibid. p. 24.

[3] See news note, Special Libraries, Jan. 1964, p. 58.

[4] Atherton and Yovich, 1962 [26], p. 22.

Kessler himself and his associates have also conducted some experiments in comparative evaluation of indexing aids derived from citation data on the one hand and from conventional subject indexing on the other. The basis for evaluation was a total of 334 papers published in The Physical Review in 1958. The study involved detailed comparison of the ways in which these papers fell into related groups according to the "analytic subject index" used by the journal's editors and according to the method of "bibliographic coupling". The essentials of the latter method are described as follows:

> "a. A single item of reference used by two papers is called one unit of coupling between them.
>
> "b. A number of papers constitute a related group G_A, if each member of the group has at least one coupling unit to a given test paper P_O.
>
> "c. The coupling strength between P_O and any member of G_A is measured by the number of coupling units (n) between them." 1/

For the 334 papers, 73 categories of the Analytic Subject Index (ASI) had been used. For the bibliographic coupling method, each of the papers was in turn considered as the test paper and groups were formed for any of the 333 other papers that shared one or more citations with it. In general, it was concluded that there was good correlation between the groupings of papers achieved by the two methods. It should be noted, however, that 44 papers fell into no groups at all on the basis of the bibliographic coupling criterion. 2/

Salton and associates at the Harvard Computation Laboratory are also concerned with the citation indexing principle as a possible basis for grouping similar documents. They are also concerned with evaluation of results so obtained by comparison with document groups obtained by subject indexing means. In the comparative experiments, data were first compiled for a closed document set of 62 items as to similarities with respect to both "citedness" and "citingness". The same items were manually indexed and similarity coefficients between these items were derived from overlappings of assigned index terms. When the two measures of similarity were compared with each other and with document associations obtained by random assignments of "citations" and "terms", the conclusions reached were as follows:

> "The similarity coefficients obtained by comparing overlapping citations for a sample document collection with overlapping, manually generated index terms are much larger than those obtained by assuming a random assignment of citations and terms to the documents; relatively large similarity coefficients are generated for nearly all documents which exhibit at least a minimum number of citations; little seems to be gained by using citation links of length greater than two; for early documents, citedness furnishes a better indication than the amount of citing, and vice versa for recent documents; for documents which can both cite and be cited, equally good indications seem to be obtained by comparing citing and cited documents." 3/

1/ Kessler, 1963 [320], p. 1, footnote.

2/ Ibid, p. 5.

3/ Salton, 1962 [520], p. III-42.

In the Salton project, tests of the value of citation links for the assignment of index terms have been made by comparing the citation pattern of an "unknown" document with those of other documents in the collection to derive a set of five "related" documents, where relatedness is decided on the basis of the magnitude of the similarity coefficients for the citation links. Any index term that appears at least twice in the set of terms previously assigned to the five related documents is then assigned to the new item. In general, approximately 50% of the terms so assigned were also assigned to the same "new" items by human indexing procedures. 1/

As we have previously noted, however, the advantages of citation indexing are likely to be most effectively applied when used as part of an array of other tools. Tukey suggests, in particular, that permutation indexes of titles, as in KWIC systems, would be of great value as "starter" and "re-check" mechanisms for the use of citation indexes. 2/ Brownson reports:

> "Consideration is now being given to the possibility of experimenting with a 'hybrid' type of index that would combine permuted titles, authors, and citation data. Such an index might be more useful than any of the individual types of indexes issued singly; and, since no human indexing judgment would be involved, it could be prepared largely by machine and issued rapidly." 3/

Williams, while at ITEK, proposed a hybrid integrated index combining listings by authors, corporate authors or author affiliations, keywords-in-context from title, and references to works cited by and to works citing an item, and she also developed a sample format for selected items from several journals in the field of philosophy. 4/

Precisely such a hybrid tool was provided with the Short Papers for the A. D. I. Annual Meeting 1963, and it was indeed issued rapidly. A brief period of only two or three weeks elapsed between receipt of many of the manuscripts and the distribution of two automatically typeset volumes. The second of these volumes contains a KWIC and an author index to these papers themselves, a bibliography and citation index to all papers referenced by them, and KWIC and author indexes to the cited papers, all computer-compiled within this time period. 5/

1/ Ibid, See also Lesk 1963, [357], p. V-8.

2/ Tukey, 1962, [611], p. 12.

3/ Brownson, 1963 [82], p. 4.

4/ T. M. Williams, private communication, dated January 4, 1962.

5/ Luhn, 1963 [376], and [377], pp. 353-382.

2.5 Machine Conversion From One Index Set to Another

A final possibility in the general area of machine compilation of indexes and machine use to improve the availability of indexes is as yet in a highly speculative stage. This is the possibility of converting from one index set to another by machine look-up procedures. In the Welch Medical Library project, mentioned earlier, use was made of punched card techniques to convert from one index arrangement to another, [1] but machine-recognizable identifiers for both arrangements were explicitly encoded in the material. In recent studies at Datatrol, however, preliminary investigations have been conducted looking toward machine lookup of index-term equivalence tables in order to convert, for example, DDC descriptors to corresponding subject headings used in the AEC vocabulary.

Hammond and Rosenborg (1962 [250] and [252]) report on the compilation of a unilateral table of "indexing equivalents" between approximately 7,000 DDC descriptors and those AEC subject headings judged by them to be identical, synonymous, or "usefully" equivalent, such as one or the other being subsumed by a broader or more generic term. Findings showed 23.8% of the terms of the DDC vocabulary presumably identical to those of AEC, 38.1% of lower generic level, 7.4% of higher generic level, and 10.9% for which no useful equivalents could be found. A sample table of indexing equivalents was prepared for DDC-to-AEC conversion, but not in the opposite direction.

Since, in general, convertibility of indexing vocabularies would be desirable wherever duplication of cataloging and indexing effort is likely to occur (that is, where two or more different documentation organizations receive at least some of the same material as inputs to their systems), the results of these preliminary studies are provocative and appear to merit the further study that is being sponsored by an Interagency Task Group on Vocabulary Study of the Committee on Scientific Information, under the Federal Council for Science and Technology.

There are many substantial difficulties, however. When applied to actual indexing of the same items by the two agencies, it was found that for 277 items indexed by both AEC and DDC (then ASTIA):

> "ASTIA used a total of 2,571 descriptors, and AEC 840 subject headings... of these, 392, or roughly half of the AEC terms, were either completely or, for all practical purpose, identical." [2]

Painter (1963 [460]) made further studies of equivalency in her investigations of duplication and consistency of subject indexing at several Government agencies. For 200 items indexed by both AEC and DDC, she found 20% DDC equivalency, 67% AEC equivalency, and 30% similarity of actual indexing. She concludes, in part:

> "In considering these solutions and the statistics revealed by the studies it should be concluded that with a maximum of only 69 percent equivalency, or convertibility, and a minimum of 28 percent, there is still a large proportion of terms which will

---

[1] Garfield, 1959 [221], p. 471.

[2] Hammond 1962 [250], p. 4.

> reconstitute some other form of retrieval. This is the proposition which is involved with the problem of generics, where a term in one system subsumes two of another --- and vice-versa. An additional problem evolves in attempting to reconcile two different subject concepts, one, the subject heading which usually has a single access point and one, the uniterm or descriptor which has multiple access through coordination. Thus the practicality of a system made up of many units supplying information indexed differently, using as a basis for retrieval a table of equivalents, is questionable." 1/

Moreover, the results of tests of inter-indexer consistency rates within the same agency were not encouraging. Thus Painter further concludes:

> "The study, in combining the results of the equivalency analysis and the consistency of indexing within each system and an equivalency of only 30 percent within the broadest system, a table of equivalents is at present of little value in either a manual or a machine system. In order to apply a table of equivalents efficiently, both a high degree of consistency and a high degree of equivalency is essential." 2/

She therefore stresses that the possibilities for conversion by machine techniques from one indexing set to an equivalent set for another vocabulary are adversely affected by the generally poor rates of inter-indexer consistency. With reference both to the Cornell Studies 3/ and to corroborative findings of her own, she states:

> "The value of equivalency studies and most particularly the table of equivalents presupposes the consistency of indexing. Convertibility between systems is thus dependent on the consistency of indexing. Without consistency, the vocabularies as units are not equal; equivalencies cannot be drawn or effectively used for convertibility." 4/

1/ Painter, 1963 [466], p. 106.

2/ Ibid, p. 14.

3/ Hammond, 1962 [250]; Hammond and Rosenborg, 1962 [252].

4/ Painter, 1963, [466], p. 109. Note that these estimates of inter-indexer consistency may be quite optimistic, as discussed on pp. 117-118 of this report.

3. INDEXES GENERATED BY MACHINE--AUTOMATIC DERIVATIVE INDEXING

We have noted, in the earlier statement of the scope of this survey, a distinction between "derivative" and "assignment" indexing. This distinction is related directly to the question: "Is what can be done by machine properly termed 'abstracting', 'indexing', or 'classifying'?" It relates also, as we have remarked, to a continuing controversy far older than any question of the introduction of machine techniques--that between "word" and "concept" indexing, between "uniterms" if selected directly from the text and "descriptors" in the sense of their being indexing terms selected so as to have "a carefully specified meaning for retrieval", 1/ to say nothing of contrasts with subject heading schemes and classification schedules.

Some of the major arguments pro and con derivative (usually word) and assignment (usually concept) indexing will be considered in a subsequent section of this report on the problems of evaluating indexing methods. Nevertheless, the present popularity of automatic derivative indexes of the KWIC type, while subject to all the disadvantages typically cited for all purely derivative indexing systems, does show the actuality of automatic indexing potentialities and may in fact hold the promise of solving some of the present-day problems of subject control.

In this section, we shall consider first the straightforward word extraction techniques used in KWIC type indexes. Possibilities for modified derivative indexing by title augmentation, manipulation of word groups and use of special clues in keyword selection are then discussed, including work by Baxendale, Luhn, and Artandi. Related research and development efforts work in automatic abstracting which lend themselves to derivation of indexing terms includes proposals and experiments by Luhn, Oswald, Edmundson, Wyllys, Doyle, and Lesk and Storm, among others. Some comments will be given on the quality of modified derivative indexing by machine. Automatic derivative indexing at the time of search, as in the natural language text searching systems of Swanson, Horty, Kehl, and Ray, and Eldridge and Dennis, will be discussed in a later section of this report. 2/

3.1 KWIC Indexes

The development of computer-generated permuted-title keyword indexes, especially in the instances of Chemical Titles and B. A. S. I. C. (Biological Abstracts-Subjects-In Context) has been hailed by some as "the miracle of the decade" and "the greatest thing to happen in chemistry since the invention of the test tube". 3/ The major reason for the optimistic enthusiasm is the speed with which the computer can produce can produce a complete index to some specific set of books, documents or papers so that publication and dissemination of the index can be prompt and thus serve as an important tool in

---

1/ Mooers, 1963 [423], p. 3.

2/ See pp. 112-116.

3/ Quoted by D. R. Baker statement in "U. S. Congress, Senate Committee on Government Operations", 1960 [615], p. 149.

maintenance of truly current awareness. For example, Herner in his 1961 review of the state-of-the-art of organizing information says:

> "I am told that the American Chemical Society has never had a more successful basic science publication. The key to the whole thing is, I believe, the extreme currency of Chemical Titles. This in turn derives from the speed and simplicity of the KWIC process." 1/

Conrad reports as follows:

> "Reception of B.A.S.I.C. ... has been so extremely enthusiastic ... that we are excited by the possibilities of producing permuted title indexes in one or more additional languages. The creation of a B.A.S.I.C. index in any language requires only that the titles be translated and punched on cards. Alphabetical arrangement, permutation and 'type-setting' is completely automated and, for 5,000 titles takes only two hours to accomplish." 2/

3.1.1 Applications of KWIC Indexing Techniques

The KWIC type process is indeed simple and straightforward. The words of the author's title are prepared for input to the computer by keystroking, either to punched cards or to punched paper tape. After being read by the computer, the text of a title is normally processed against a "stop list" to eliminate from further processing the more common words, such as "the", "and", prepositions, and the like, and words so general as to be insignificant for indexing purposes, such as "demonstration", "typical", "measurements", "steps", and the like. The remaining presumably "significant" or "key" words are then, in effect, taken one at a time to an indexing position or window, where they are sorted in alphabetical order. The result is a listing of each such word together with its surrounding context, out to the limit of the line or lines permitted in a given format. As each keyword is processed, the title itself is moved over so that the next keyword occupies the indexing position, and this process is repeated until the entire title has thus been cyclically permuted.

A number of formats are available in which the length of the line, the position of the indexing window, and the extent of "wrap-around" (bringing the end of a title in at the beginning of a line to fill space that would otherwise be left blank) are major variables. Current examples of KWIC type indexing output are shown in Figures 2 through 7. Usually, the indexing window is located at or near the center of the line with several extra spaces to the immediate left or with other devices such as the shading of B.A.S.I.C. to aid the searcher in scanning down the keywords listed. This is

---

1/ Herner, 1962, [266], p. 10.

2/ Conrad, 1962 [137], p. 378A.

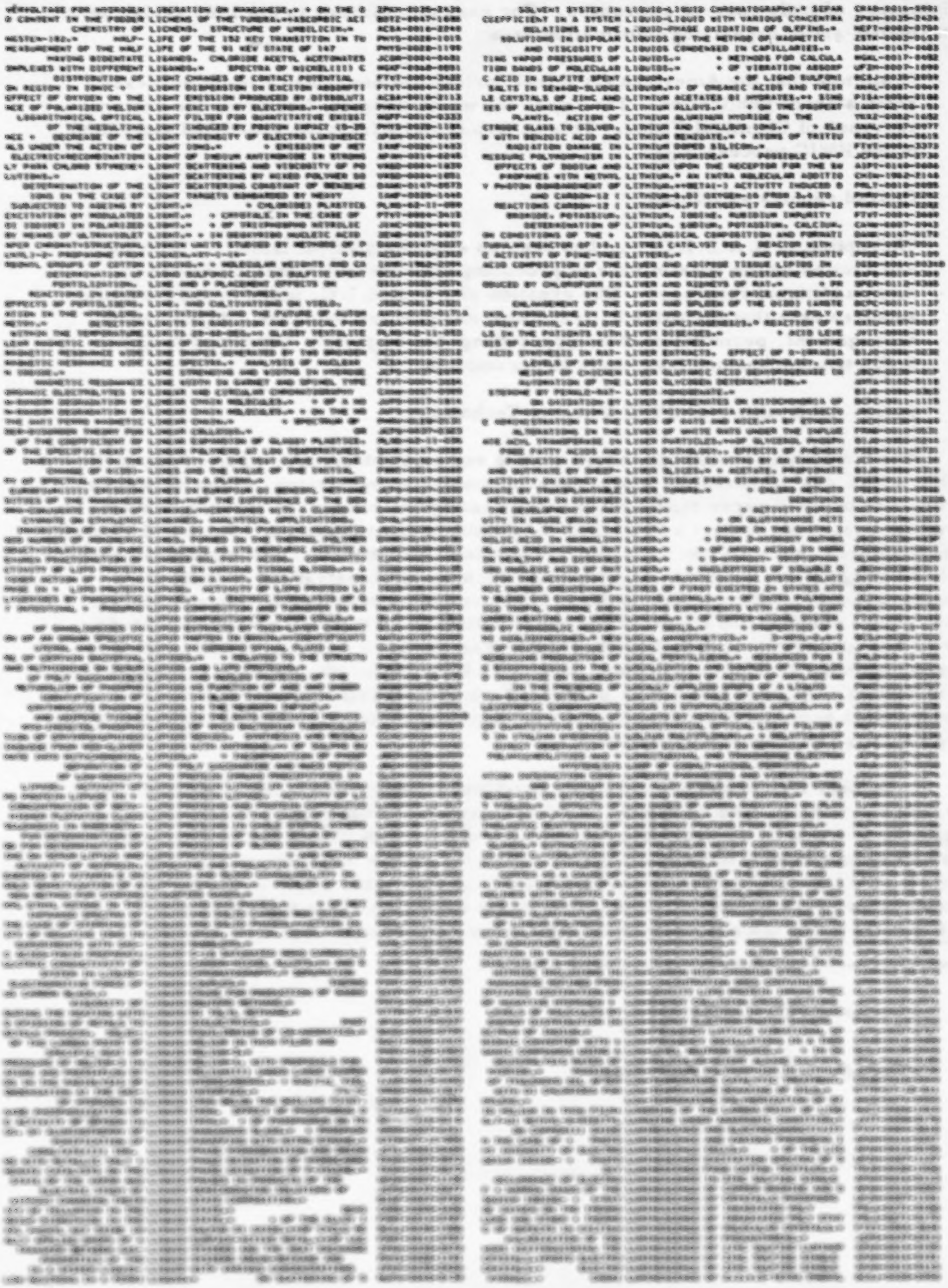
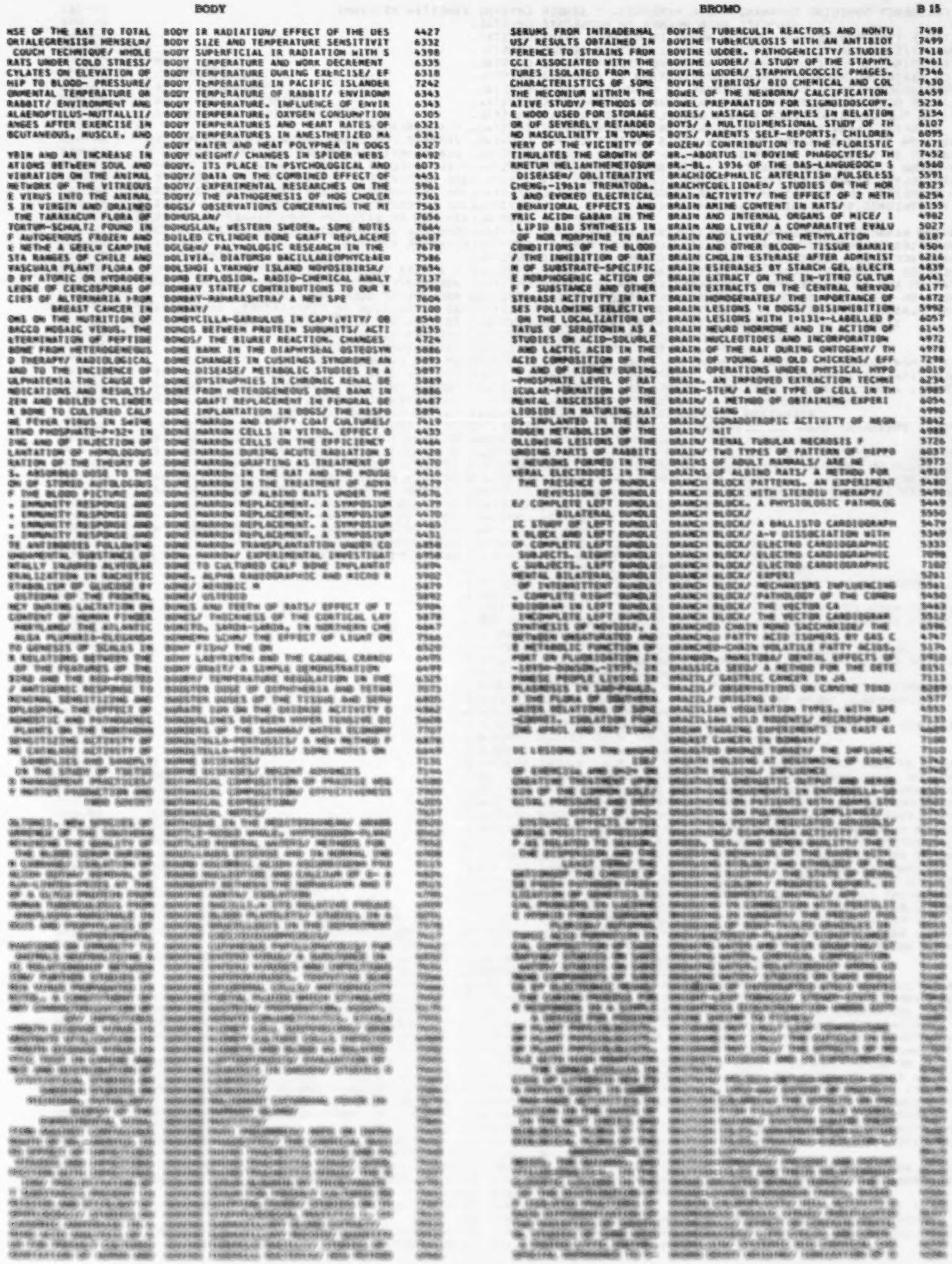


Figure 2. Sample Page, Standard IBM Format

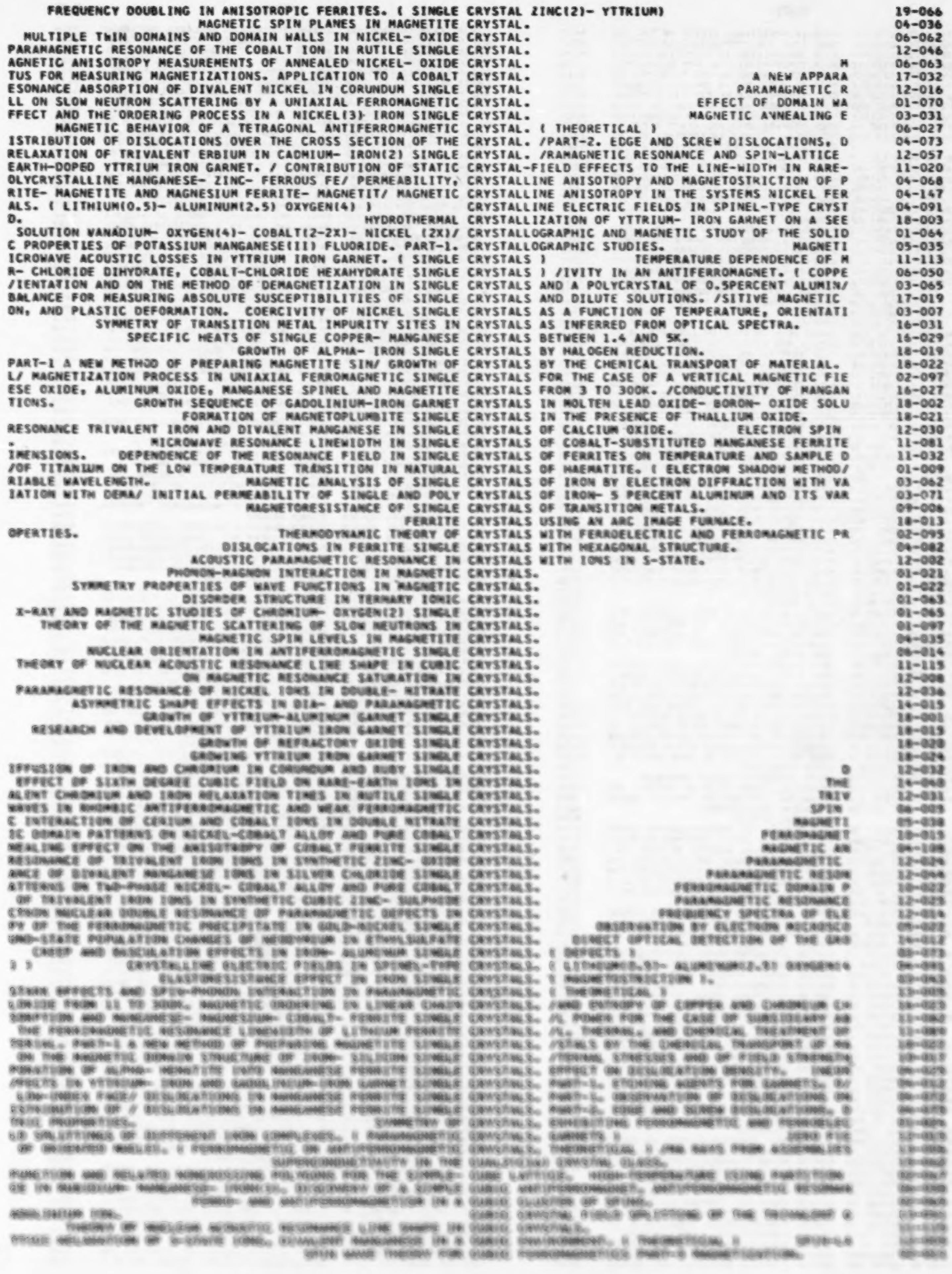


BODY BROMO B 15

NSE OF THE RAT TO TOTAL	BODY IR RADIATION/ EFFECT OF THE DES	4427
ORTALEGRENSIS= HENSELØ/	BODY SIZE AND TEMPERATURE SENSITIVIT	6332
COUCH TECHNIQUE/ WHOLE	BODY SUPERFICIAL IR RADIATION WITH S	4398
RATS UNDER COLD STRESS/	BODY TEMPERATURE AND WORK DECREMENT	6335
CYLATES ON ELEVATION OF	BODY TEMPERATURE DURING EXERCISE/ EF	6318
HIP TO BLOOD- PRESSURE/	BODY TEMPERATURE IN PACIFIC ISLANDER	7242
ONMENTAL TEMPERATURE ON	BODY TEMPERATURE OF RABBIT/ ENVIRONM	6343
RABBIT/ ENVIRONMENT AND	BODY TEMPERATURE. INFLUENCE OF ENVIR	6343
ALAENOPTILUS-NUTTALLII/	BODY TEMPERATURE, OXYGEN CONSUMPTION	6305
ANGES AFTER EXERCISE IN	BODY TEMPERATURES AND HEART RATES OF	6321
BCUTANEOUS, MUSCLE, AND	BODY TEMPERATURES IN ANESTHETIZED MA	6341
/	BODY WATER AND HEAT POLYPNEA IN DOGS	6327
VRIN AND AN INCREASE IN	BODY WEIGHT/ CHANGES IN SPIDER WEBS	8492
ATIONS BETWEEN SOUL AND	BODY. ITS PLACE IN PSYCHOLOGICAL AND	6073
VIBRATION ON THE ANIMAL	BODY/ DATA ON THE COMBINED EFFECT OF	6451
NETWORK OF THE VITREOUS	BODY/ EXPERIMENTAL RESEARCHES ON THE	5961
E VIRUS INTO THE ANIMAL	BODY/ THE PATHOGENESIS OF HOG CHOLER	7361
S IN VIRGIN AND DRAINED	BOGS/ OBSERVATIONS CONCERNING THE MI	7563
THE TARAXACUM FLORA OF	BOHUSLAN/	7656
TORTUM-SCHULTZ FOUND IN	BOHUSLAN, WESTERN SWEDEN. SOME NOTES	7644
F AUTOGENOUS FROZEN AND	BOILED CYLINDER BONE GRAFT REPLACEME	6487
E NETHE A GEEL= CAMPINE	BOLGEN/ PALYNOLOGIC RESEARCH IN THE	7678
STA RANGES OF CHILE AND	BOLIVIA. DIATOMS= BACILLARIOPHYCEAE=	7586
VASCULAR PLANT FLORA OF	BOLSHOI LYAKHOV ISLAND NOVOSIBIRSK/	7654
D BY ATOMIC OR HYDROGEN	BOMB EXPLOSION. RADIO-CHEMICAL ANALY	7137
LEDGE OF CERCOSPORAE OF	BOMBAY STATE/ CONTRIBUTIONS TO OUR K	7598
CIES OF ALTERNARIA FROM	BOMBAY-MAHARASHTRA/ A NEW SPE	7604
BREAST CANCER IN	BOMBAY/	7100
ONS ON THE NUTRITION OF	BOMBYCILLA-GARRULUS IN CAPTIVITY/ OB	8540
BACCO MOSAIC VIRUS. THE	BONDS BETWEEN PROTEIN SUBUNITS/ ACTI	8155
ETERMINATION OF PEPTIDE	BONDS/ THE BIURET REACTION. CHANGES	4724
BONE FROM HETEROGENEOUS	BONE BANK IN THE DIAPHYSEAL OSTEOSYN	5086

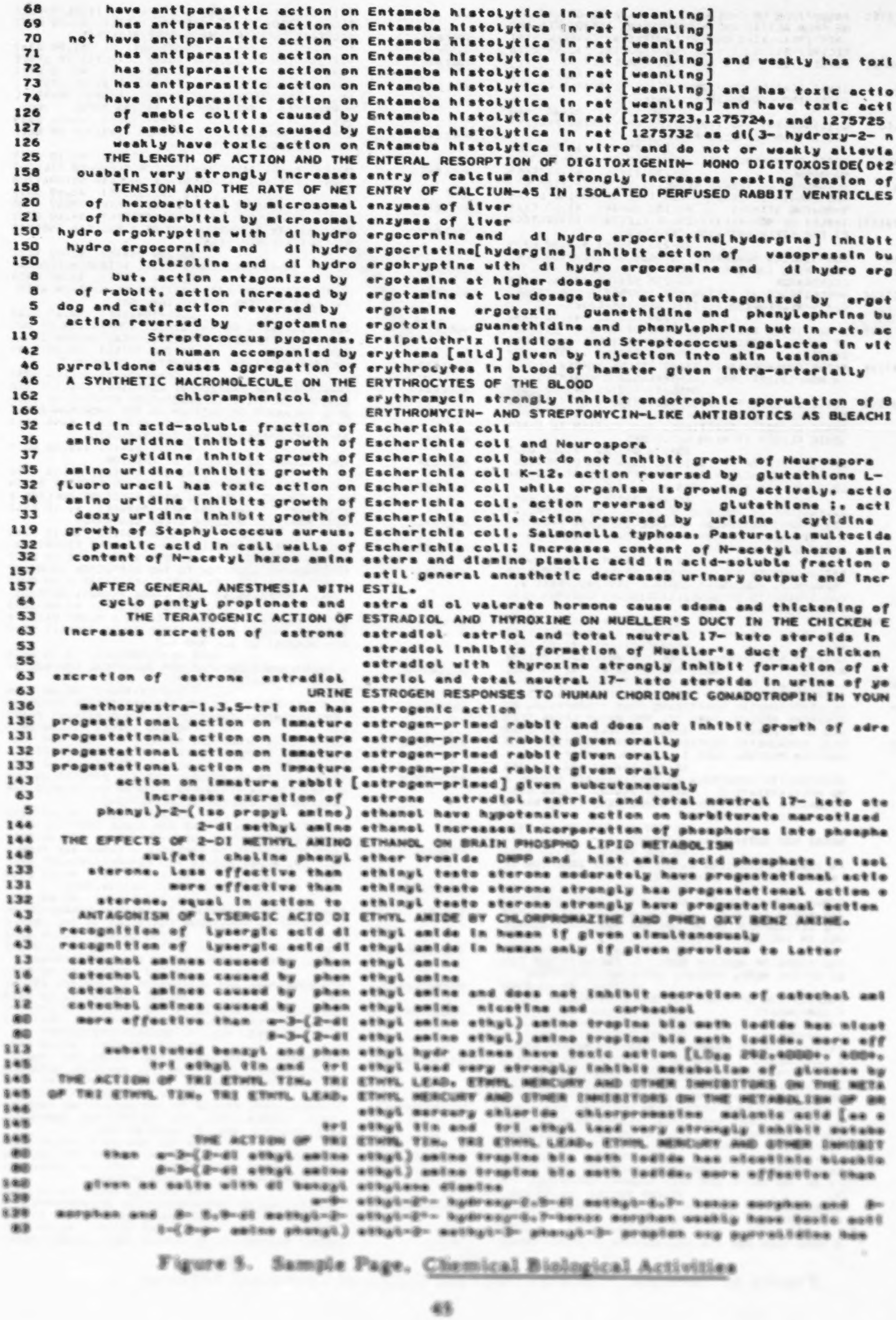
SERUMS FROM INTRADERMAL	BOVINE TUBERCULIN REACTORS AND NONTU	7498
US/ RESULTS OBTAINED IN	BOVINE TUBERCULOSIS WITH AN ANTIBIOT	7499
FERENCE TO STRAINS FROM	BOVINE UDDER. PATHOGENICITY/ STUDIES	7418
CCI ASSOCIATED WITH THE	BOVINE UDDER/ A STUDY OF THE STAPHYL	7461
TURES ISOLATED FROM THE	BOVINE UDDER/ STAPHYLOCOCCIC PHAGES.	7346
CHARACTERISTICS OF SOME	BOVINE VIBRIOS/ BIO CHEMICAL AND COL	7430
THE MECONIUM WITHIN THE	BOWEL OF THE NEWBORN/ CALCIFICATION	6459
ATIVE STUDY/ METHODS OF	BOWEL PREPARATION FOR SIGMOIDOSCOPY.	5236
E WOOD USED FOR STORAGE	BOXES/ WASTAGE OF APPLES IN RELATION	5154
OR OF SEVERELY RETARDED	BOYS/ A MULTIDIMENSIONAL STUDY OF TH	6107
ND MASCULINITY IN YOUNG	BOYS/ PARENTS SELF-REPORTS, CHILDREN	6095
VERY OF THE VICINITY OF	BOZEN/ CONTRIBUTION TO THE FLORISTIC	7671
TIMULATES THE GROWTH OF	BR.-ABORTUS IN BOVINE PHAGOCYTES/ TH	7452
RMETUM HELIANTHEMETOSUM	BR.-BL. 1936 OF THE BAS-LANGUEDOC= S	4560
DISEASE=/ OBLITERATIVE	BRACHIOCEPHALIC ARTERITIS= PULSELESS	5591
CHENG.-1961= TREMATODA.	BRACHYCOELIIDAE=/ STUDIES ON THE MOR	8323
S AND EVOKED ELECTRICAL	BRAIN ACTIVITY/ THE EFFECT OF 2 METH	6254
BEHAVIORAL EFFECTS AND	BRAIN AMINE CONTENT IN RATS/	6155
YRIC ACID= GABA= IN THE	BRAIN AND INTERNAL ORGANS OF MICE/ I	4982
LIPID BIO SYNTHESIS IN	BRAIN AND LIVER/ A COMPARATIVE EVALU	5027
OF NOR MORPHINE IN RAT	BRAIN AND LIVER/ THE METHYLATION	6187
CONDITIONS OF THE BLOOD	BRAIN AND OTHER BLOOD- TISSUE BARRIE	4504
/ THE INHIBITION OF RAT	BRAIN CHOLIN ESTERASE AFTER ADMINIST	6216
N OF SUBSTRATE-SPECIFIC	BRAIN ESTERASES BY STARCH GEL ELECTR	4839
E MORPHOGENIC ACTION OF	BRAIN EXTRACT ON THE IN-VITRO CULTUR	6441
P P SUBSTANCE AND OTHER	BRAIN EXTRACTS ON THE CENTRAL NERVOU	6177
STERASE ACTIVITY IN RAT	BRAIN HOMOGENATES/ THE IMPORTANCE OF	4872
SES FOLLOWING SELECTIVE	BRAIN LESIONS IN DOGS/ DISINHIBITION	5992
ON THE LOCALIZATION OF	BRAIN LESIONS WITH I=131=-LABELLED P	6057
TATUS OF SEROTONIN AS A	BRAIN NEURO HORMONE AND IN ACTION OF	6145
STUDIES ON ACID-SOLUBLE	BRAIN NUCLEOTIDES AND INCORPORATION	4972
AND LACTIC ACID IN THE	BRAIN OF THE RAT DURING ONTOGENY/ TH	4978

Figure 3. Sample Page. B.A.S.I.C.



```
     FREQUENCY DOUBLING IN ANISOTROPIC FERRITES. ( SINGLE CRYSTAL ZINC(2)- YTTRIUM)                           19-066
                        MAGNETIC SPIN PLANES IN MAGNETITE CRYSTAL.                                            04-036
  MULTIPLE THIN DOMAINS AND DOMAIN WALLS IN NICKEL- OXIDE CRYSTAL.                                            06-062
PARAMAGNETIC RESONANCE OF THE COBALT ION IN RUTILE SINGLE CRYSTAL.                                            12-046
AGNETIC ANISOTROPY MEASUREMENTS OF ANNEALED NICKEL- OXIDE CRYSTAL.                                       M    06-063
TUS FOR MEASURING MAGNETIZATIONS. APPLICATION TO A COBALT CRYSTAL.                            A NEW APPARA    17-032
ESONANCE ABSORPTION OF DIVALENT NICKEL IN CORUNDUM SINGLE CRYSTAL.                          PARAMAGNETIC R    12-016
LL ON SLOW NEUTRON SCATTERING BY A UNIAXIAL FERROMAGNETIC CRYSTAL.                     EFFECT OF DOMAIN WA    01-070
FFECT AND THE ORDERING PROCESS IN A NICKEL(3)- IRON SINGLE CRYSTAL.                   MAGNETIC ANNEALING E    03-031
       MAGNETIC BEHAVIOR OF A TETRAGONAL ANTIFERROMAGNETIC CRYSTAL. ( THEORETICAL )                           06-027
ISTRIBUTION OF DISLOCATIONS OVER THE CROSS SECTION OF THE CRYSTAL. /PART-2. EDGE AND SCREW DISLOCATIONS, D    04-073
RELAXATION OF TRIVALENT ERBIUM IN CADMIUM- IRON(2) SINGLE CRYSTAL. /RAMAGNETIC RESONANCE AND SPIN-LATTICE     12-057
EARTH-DOPED YTTRIUM IRON GARNET. / CONTRIBUTION OF STATIC CRYSTAL-FIELD EFFECTS TO THE LINE-WIDTH IN RARE-    11-020
OLYCRYSTALLINE MANGANESE- ZINC- FERROUS FE/ PERMEABILITY, CRYSTALLINE ANISOTROPY AND MAGNETOSTRICTION OF P    04-068
RITE- MAGNETITE AND MAGNESIUM FERRITE- MAGNETIT/ MAGNETIC CRYSTALLINE ANISOTROPY IN THE SYSTEMS NICKEL FER    04-147
ALS. ( LITHIUM(0.5)- ALUMINUM(2.5) OXYGEN(4) )            CRYSTALLINE ELECTRIC FIELDS IN SPINEL-TYPE CRYST    04-091
D.                                           HYDROTHERMAL CRYSTALLIZATION OF YTTRIUM- IRON GARNET ON A SEE    18-003
 SOLUTION VANADIUM- OXYGEN(4)- COBALT(2-2X)- NICKEL (2X)/ CRYSTALLOGRAPHIC AND MAGNETIC STUDY OF THE SOLID    01-064
C PROPERTIES OF POTASSIUM MANGANESE(III) FLUORIDE. PART-1. CRYSTALLOGRAPHIC STUDIES.               MAGNETI    05-035
ICROWAVE ACOUSTIC LOSSES IN YTTRIUM IRON GARNET. ( SINGLE CRYSTALS )           TEMPERATURE DEPENDENCE OF M    11-113
R- CHLORIDE DIHYDRATE, COBALT-CHLORIDE HEXAHYDRATE SINGLE CRYSTALS ) /IVITY IN AN ANTIFERROMAGNET. ( COPPE    06-050
/IENTATION AND ON THE METHOD OF DEMAGNETIZATION IN SINGLE CRYSTALS AND A POLYCRYSTAL OF 0.5PERCENT ALUMIN/    03-065
BALANCE FOR MEASURING ABSOLUTE SUSCEPTIBILITIES OF SINGLE CRYSTALS AND DILUTE SOLUTIONS. /SITIVE MAGNETIC     17-019
ON, AND PLASTIC DEFORMATION.  COERCIVITY OF NICKEL SINGLE CRYSTALS AS A FUNCTION OF TEMPERATURE, ORIENTATI    03-007
           SYMMETRY OF TRANSITION METAL IMPURITY SITES IN CRYSTALS AS INFERRED FROM OPTICAL SPECTRA.           16-031
               SPECIFIC HEATS OF SINGLE COPPER- MANGANESE CRYSTALS BETWEEN 1.4 AND 5K.                        16-029
                             GROWTH OF ALPHA- IRON SINGLE CRYSTALS BY HALOGEN REDUCTION.                      18-019
PART-1 A NEW METHOD OF PREPARING MAGNETITE SIN/ GROWTH OF CRYSTALS BY THE CHEMICAL TRANSPORT OF MATERIAL.     18-022
L/ MAGNETIZATION PROCESS IN UNIAXIAL FERROMAGNETIC SINGLE CRYSTALS FOR THE CASE OF A VERTICAL MAGNETIC FIE    02-097
ESE OXIDE, ALUMINUM OXIDE, MANGANESE SPINEL AND MAGNETITE CRYSTALS FROM 3 TO 300K. /CONDUCTIVITY OF MANGAN    16-027
TIONS.          GROWTH SEQUENCE OF GADOLINIUM-IRON GARNET CRYSTALS IN MOLTEN LEAD OXIDE- BORON- OXIDE SOLU    18-002
                      FORMATION OF MAGNETOPLUMBITE SINGLE CRYSTALS IN THE PRESENCE OF THALLIUM OXIDE.         18-021
RESONANCE TRIVALENT IRON AND DIVALENT MANGANESE IN SINGLE CRYSTALS OF CALCIUM OXIDE.        ELECTRON SPIN     12-030
.                 MICROWAVE RESONANCE LINEWIDTH IN SINGLE CRYSTALS OF COBALT-SUBSTITUTED MANGANESE FERRITE    11-081
IMENSIONS.    DEPENDENCE OF THE RESONANCE FIELD IN SINGLE CRYSTALS OF FERRITES ON TEMPERATURE AND SAMPLE D    11-032
/OF TITANIUM ON THE LOW TEMPERATURE TRANSITION IN NATURAL CRYSTALS OF HAEMATITE. ( ELECTRON SHADOW METHOD/    01-009
RIABLE WAVELENGTH.            MAGNETIC ANALYSIS OF SINGLE CRYSTALS OF IRON BY ELECTRON DIFFRACTION WITH VA    03-062
IATION WITH DEMA/ INITIAL PERMEABILITY OF SINGLE AND POLY CRYSTALS OF IRON- 5 PERCENT ALUMINUM AND ITS VAR    03-071
                              MAGNETORESISTANCE OF SINGLE CRYSTALS OF TRANSITION METALS.                      09-006
                                                  FERRITE CRYSTALS USING AN ARC IMAGE FURNACE.                18-013
OPERTIES.                         THERMODYNAMIC THEORY OF CRYSTALS WITH FERROELECTRIC AND FERROMAGNETIC PR    02-095
                           DISLOCATIONS IN FERRITE SINGLE CRYSTALS WITH HEXAGONAL STRUCTURE.                  04-082
                       ACOUSTIC PARAMAGNETIC RESONANCE IN CRYSTALS WITH IONS IN S-STATE.                      12-002
                    PHONON-MAGNON INTERACTION IN MAGNETIC CRYSTALS.                                           01-021
        SYMMETRY PROPERTIES OF WAVE FUNCTIONS IN MAGNETIC CRYSTALS.                                           01-022
                      DISORDER STRUCTURE IN TERNARY IONIC CRYSTALS.                                           01-063
 X-RAY AND MAGNETIC STUDIES OF CHROMIUM- OXYGEN(2) SINGLE CRYSTALS.                                           01-065
    THEORY OF THE MAGNETIC SCATTERING OF SLOW NEUTRONS IN CRYSTALS.                                           01-097
                        MAGNETIC SPIN LEVELS IN MAGNETITE CRYSTALS.                                           04-035
          NUCLEAR ORIENTATION IN ANTIFERROMAGNETIC SINGLE CRYSTALS.                                           06-014
 THEORY OF NUCLEAR ACOUSTIC RESONANCE LINE SHAPE IN CUBIC CRYSTALS.                                           11-115
                      ON MAGNETIC RESONANCE SATURATION IN CRYSTALS.                                           12-008
 PARAMAGNETIC RESONANCE OF NICKEL IONS IN DOUBLE- NITRATE CRYSTALS.                                           12-036
        ASYMMETRIC SHAPE EFFECTS IN DIA- AND PARAMAGNETIC CRYSTALS.                                           14-015
                 GROWTH OF YTTRIUM-ALUMINUM GARNET SINGLE CRYSTALS.                                           18-001
   RESEARCH AND DEVELOPMENT OF YTTRIUM IRON GARNET SINGLE CRYSTALS.                                           18-015
                        GROWTH OF REFRACTORY OXIDE SINGLE CRYSTALS.                                           18-020
                       GROWING YTTRIUM IRON GARNET SINGLE CRYSTALS.                                           18-024
IFFUSION OF IRON AND CHROMIUM IN CORUNDUM AND RUBY SINGLE CRYSTALS.                                      D    12-032
```

Figure 4. Sample, Bell Laboratories Format

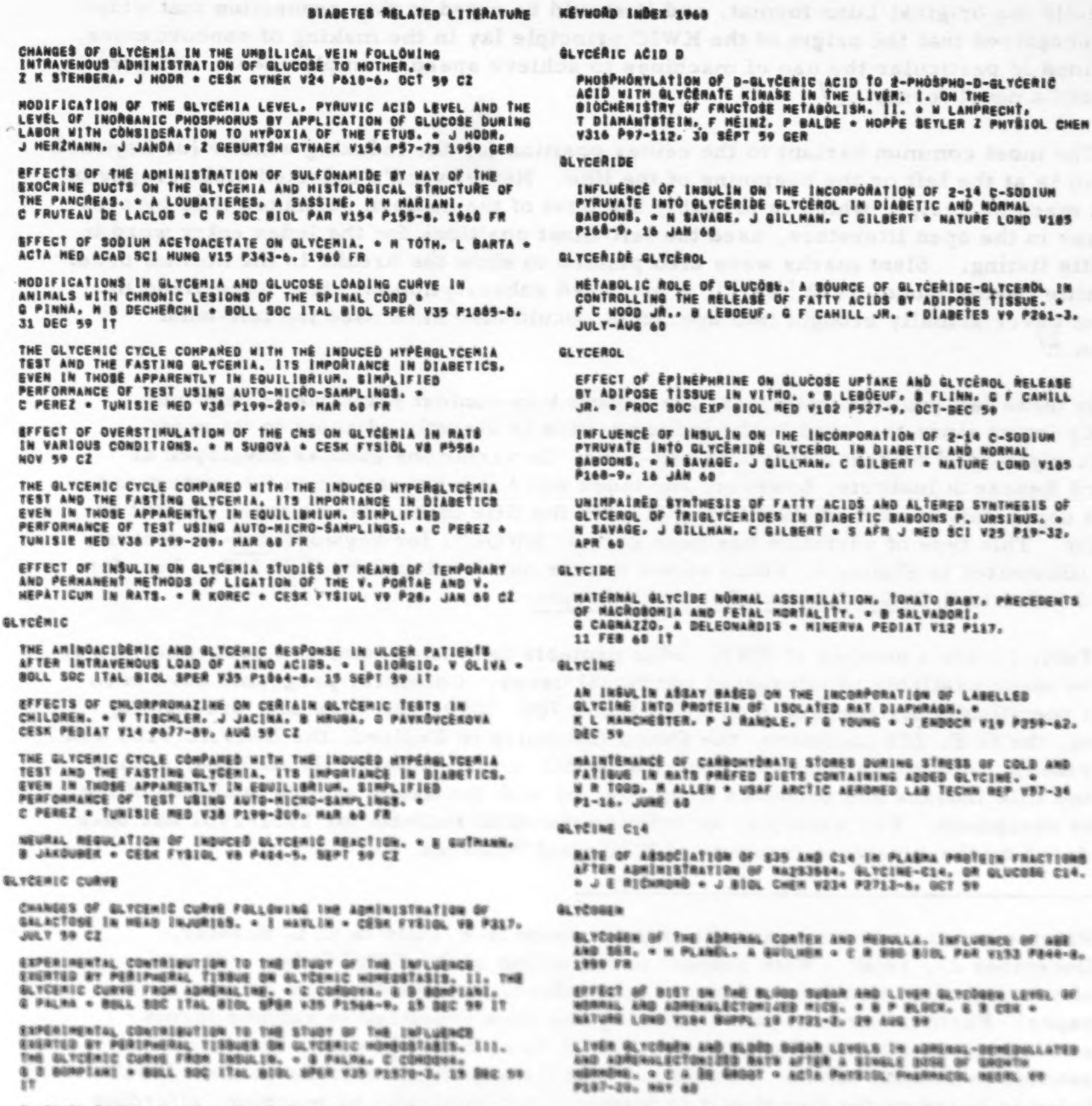


```
 68         have antiparasitic action on Entameba histolytica in rat [weanling]
 69          has antiparasitic action on Entameba histolytica in rat [weanling]
 70     not have antiparasitic action on Entameba histolytica in rat [weanling]
 71          has antiparasitic action on Entameba histolytica in rat [weanling] and weakly has toxi
 72          has antiparasitic action on Entameba histolytica in rat [weanling]
 73          has antiparasitic action on Entameba histolytica in rat [weanling] and has toxic actio
 74         have antiparasitic action on Entameba histolytica in rat [weanling] and have toxic acti
126          of amebic colitis caused by Entameba histolytica in rat [1275723,1275724, and 1275725
127          of amebic colitis caused by Entameba histolytica in rat [1275732 as di(3- hydroxy-2- n
126          weakly have toxic action on Entameba histolytica in vitro and do not or weakly allevia
 25        THE LENGTH OF ACTION AND THE ENTERAL RESORPTION OF DIGITOXIGENIN- MONO DIGITOXOSIDE(Dt2
158     ouabain very strongly increases entry of calcium and strongly increases resting tension of
158          TENSION AND THE RATE OF NET ENTRY OF CALCIUN-45 IN ISOLATED PERFUSED RABBIT VENTRICLES
 20       of  hexobarbital by microsomal enzymes of liver
 21       of  hexobarbital by microsomal enzymes of liver
150   hydro ergokryptine with  di hydro ergocornine and   di hydro ergocristine[hydergine] inhibit
150    hydro ergocornine and   di hydro ergocristine[hydergine] inhibit action of   vasopressin bu
150             tolazoline and  di hydro ergokryptine with  di hydro ergocornine and   di hydro erg
  8        but, action antagonized by  ergotamine at higher dosage
  8     of rabbit, action increased by  ergotamine at low dosage but, action antagonized by  erget
  5   dog and cat; action reversed by   ergotamine  ergotoxin   guanethidine and  phenylephrine bu
  5    action reversed by   ergotamine  ergotoxin   guanethidine and  phenylephrine but in rat, ac
119              Streptococcus pyogenes, Ersipelothrix insidiosa and Streptococcus agalactae in vit
 42              in human accompanied by erythema [mild] given by injection into skin lesions
 46   pyrrolidone causes aggregation of erythrocytes in blood of hamster given intra-arterially
 46    A SYNTHETIC MACROMOLECULE ON THE ERYTHROCYTES OF THE BLOOD
162                 chloramphenicol and  erythromycin strongly inhibit endotrophic sporulation of B
166                                      ERYTHROMYCIN- AND STREPTOMYCIN-LIKE ANTIBIOTICS AS BLEACHI
 32    acid in acid-soluble fraction of Escherichia coli
 36    amino uridine inhibits growth of Escherichia coli and Neurospora
 37          cytidine inhibit growth of Escherichia coli but do not inhibit growth of Neurospora
 35    amino uridine inhibits growth of Escherichia coli K-12, action reversed by  glutathione L-
 32   fluoro uracil has toxic action on Escherichia coli while organism is growing actively, actio
 34    amino uridine inhibits growth of Escherichia coli, action reversed by   glutathione :, acti
 33     deoxy uridine inhibit growth of Escherichia coli, action reversed by  uridine   cytidine
119    growth of Staphylococcus aureus, Escherichia coli, Salmonella typhosa, Pasturella multocida
 32       pimelic acid in cell walls of Escherichia coli; increases content of N-acetyl hexos amin
 32     content of N-acetyl hexos amine esters and diamino pimelic acid in acid-soluble fraction o
157                                      estil general anesthetic decreases urinary output and incr
157        AFTER GENERAL ANESTHESIA WITH ESTIL.
 64         cyclo pentyl propionate and  estra di ol valerate hormone cause edema and thickening of
 53             THE TERATOGENIC ACTION OF ESTRADIOL AND THYROXINE ON MUELLER'S DUCT IN THE CHICKEN E
 63    increases excretion of  estrone  estradiol  estriol and total neutral 17- keto steroids in
 53                                      estradiol inhibits formation of Mueller's duct of chicken
 55                                      estradiol with  thyroxine strongly inhibit formation of st
 63   excretion of  estrone  estradiol  estriol and total neutral 17- keto steroids in urine of yo
 63                               URINE ESTROGEN RESPONSES TO HUMAN CHORIONIC GONADOTROPIN IN YOUN
136       methoxyestra-1,3,5-tri ene has estrogenic action
135   progestational action on immature estrogen-primed rabbit and does not inhibit growth of adre
131   progestational action on immature estrogen-primed rabbit given orally
132   progestational action on immature estrogen-primed rabbit given orally
133   progestational action on immature estrogen-primed rabbit given orally
143           action on immature rabbit [estrogen-primed] given subcutaneously
 63              increases excretion of  estrone  estradiol  estriol and total neutral 17- keto ste
  5        phenyl)-2-(iso propyl amine) ethanol have hypotensive action on barbiturate narcotized
144                    2-di methyl amino ethanol increases incorporation of phosphorus into phospho
144    THE EFFECTS OF 2-DI METHYL AMINO ETHANOL ON BRAIN PHOSPHO LIPID METABOLISM
148             sulfate  choline phenyl ether bromide  DMPP and  hist amine acid phosphate in isol
133       sterone, less effective than  ethinyl testo sterone moderately have progestational actio
131                more effective than  ethinyl testo sterone strongly has progestational action o
132       sterone, equal in action to  ethinyl testo sterone strongly have progestational action
 43     ANTAGONISM OF LYSERGIC ACID DI ETHYL AMIDE BY CHLORPROMAZINE AND PHEN OXY BENZ AMINE.
 44   recognition of  lysergic acid di ethyl amide in human if given simultaneously
 43   recognition of  lysergic acid di ethyl amide in human only if given previous to latter
 13     catechol amines caused by  phen ethyl amine
 16     catechol amines caused by  phen ethyl amine
 14     catechol amines caused by  phen ethyl amine and does not inhibit secretion of catechol ami
 12     catechol amines caused by  phen ethyl amine  nicotine and   carbachol
 80      more effective than  a-3-(2-di ethyl amino ethyl) amino tropine bis meth iodide has nicot
 80                          B-3-(2-di ethyl amino ethyl) amino tropine bis meth iodide, more eff
113          substituted benzyl and phen ethyl hydr azines have toxic action [LD50 292,4000+, 400+,
145                 tri ethyl tin and  tri ethyl lead very strongly inhibit metabolism of  glucose by
145    THE ACTION OF TRI ETHYL TIN, TRI ETHYL LEAD, ETHYL MERCURY AND OTHER INHIBITORS ON THE META
145   OF TRI ETHYL TIN, TRI ETHYL LEAD, ETHYL MERCURY AND OTHER INHIBITORS ON THE METABOLISM OF BR
146                                      ethyl mercury chloride  chlorpromazine  malonic acid [as s
145                                  tri ethyl tin and  tri ethyl lead very strongly inhibit metabo
145                     THE ACTION OF TRI ETHYL TIN, TRI ETHYL LEAD, ETHYL MERCURY AND OTHER INHIBIT
 80          than  a-3-(2-di ethyl amino ethyl) amino tropine bis meth iodide has nicotinic blockin
 80                 B-3-(2-di ethyl amino ethyl) amino tropine bis meth iodide, more effective than
162      given as salts with di benzyl ethylene diamine
139                                  a-5- ethyl-2'- hydroxy-2,9-di methyl-6,7- benzo morphan and  B-
139   morphan and  B- 5,9-di methyl-2- ethyl-2'- hydroxy-6,7-benzo morphan weakly have toxic acti
 83                 1-(2-p- amino phenyl) ethyl-2- methyl-3- phenyl-3- propion oxy pyrrolidine has
```

Figure 5. Sample Page, Chemical Biological Activities

```
NON-IRRADIATED  ABSORPTION OF D-GLUCOSE BY SEGMENTS OF INTESTI
                NE FROM ACTIVE AND HIBERNATING, IRRADIATED AND
                 NON-IRRADIATED GROUND SQUIRRELS, CITELLUS TRI
                DECEMLINEATUS    NASA N63-11002(K)  $2.60 0726
NON-ISOTHERMAL  CORRELATIONS IN A NON-ISOTHERMAL PLASMA
                                   AD-290 053(K)  $1.10 0196
NON-LINEAR      INVESTIGATION OF MICROWAVE NON-LINEAR EFFECTS
                UTILIZING FERROMAGNETIC MATERIALS
                                   AD-290 572(K)  $2.60 0487
NON-METALLIC    BIBLIOGRAPHY AND TABULATION OF DAMPING PROPERT
                IES OF NON-METALLIC MATERIALS
                                   AD-289 856(K)  $3.00 0502
NON-MILITARY    NOTES ON NON-MILITARY MEASURES IN CONTROL OF I
                NSURGENCY          AD-290 237(K)  $1.60 0696
NON-MOVING      JUDGMENTS OF VISUAL VELOCITY AS A FUNCTION OF
                THE LENGTH OF OBSERVATION TIME OF MOVING OR NO
                N-MOVING STIMULI    PB 162 549(K)  $1.60 0125
NON-RELATIVISTI TABLES OF NON-RELATIVISTIC ELECTRON TRAJECTORI
                ES FOR FIELD EMISSION CATHCDES
                                   AD-290 696(K) $14.50 0239
NON-SIMILAR     NON-SIMILAR NUMERICAL METHODS OF SOLUTION FOR
                ELECTRODE BOUNDARY LAYERS IN A CROSSED FIELD A
                CCELERATOR         AD-290 525(K)  $5.60 0185
NONDESTRUCTIVE  NONDESTRUCTIVE SYSTEM FOR INSPECTION OF FIBER
                GLASS-REINFORCED PLASTIC MISSILE CASES
                                   AD-289 825(K)  $1.60 0632
NONDESTRUCTIVE  X-RAY IMAGE SYSTEM FOR NONDESTRUCTIVE TESTING
                OF SOLID PROPELLANT MISSILE CASE WALLS AND WEL
                DMENTS             AD-289 821(K)  $3.60 0637
NONDISSIPATIVE  MAGNETOHYDRODYNAMIC STABILITY OF VORTEX FLOW -
                 A NONDISSIPATIVE, INCOMPRESSIBLE ANALYSIS
                                  ORNL-TM-402(K)  $3.60 0615
NONEQUILIBRIUM  SCALE EFFECTS FOR NONEQUILIBRIUM CONVECTIVE HE
                AT TRANSFER WITH SIMULTANEOUS GAS PHASE AND SU
                RFACE CHEMICAL REACTIONS. APPLICATION TO HYPER
                SONIC FLIGHT AT HIGH ALTITUDES
                                   AD-291 032(K)  $1.60 0025
NONLINEAR       APPLICATION OF VARIATIONAL EQUATION OF MOTION
                TO THE NONLINEAR VIBRATION ANALYSIS OF HOMOGEN
                EOUS AND LAYERED PLATES AND SHELLS
                                   AD-289 868(K)  $2.60 0667
NONLINEAR       EXTENSIONS IN THE SYNTHESIS OF TIME OPTIMAL OR
                 BANG-BANG NONLINEAR CONTROL SYSTEMS. PART I.
                THE SYNTHESIS OF QUASI-STATIONARY OPTIMUM NONL
                INEAR CONTROL SYSTEMS
                                    PB 162 547(K)  $4.60 0235
NONLINEAR       EXTENSIONS IN THE SYNTHESIS OF TIME OPTIMAL OR
                 BANG-BANG NONLINEAR CONTROL SYSTEMS. PART I.
                THE SYNTHESIS OF QUASI-STATIONARY OPTIMUM NONL
                INEAR CONTROL SYSTEMS
                                    PB 162 547(K)  $4.60 0235
NONLINEAR       NONLINEAR FLEXURAL VIBRATIONS OF SANDWICH PLAT
                ES                 AD-289 871(K)  $2.60 0669
NONLINEAR       OPTIMUM NONLINEAR CONTROL FOR ARBITRARY DISTUR
                BANCES           NASA N62-15890(K)  $2.60 0682
NONRECURRENT    A TECHNIQUE FOR NARROW-BAND TELEMETRY OF NONRE
                CURRENT PULSES     AD-290 697(K)  $2.60 0577
NONUNIFORM      ELECTROMAGNETIC SCATTERING FROM A SPHERICAL NO
                NUNIFORM MEDIUM. PART II. THE RADAR CROSS SECT
                ION OF A FLARE     AD-289 615(K)  $2.60 0747
NONUNIFORM      ELECTROMAGNETIC SCATTERING FROM ASPHERICAL NON
                UNIFORM MEDIUM. PART I. GENERAL THEORY
                                   AD-289 614(K)  $2.60 0748
NORMAL          PROBABILITY INTEGRALS OF MULTIVARIATE NORMAL A
                ND MULTIVARIATE-T  AD-290 746(K)  $8.60 0760
NORMAL          RESONANCE ABSORPTION OF GAMMA-RAYS IN NORMAL A
                ND SUPERCONDUCTING TIN
                                   AD-289 844(K)  $3.60 0826
NORMS           NORMS FOR ARTIFICIAL LIGHTING
                                   AD-290 355(K)  $1.10 0734
NORTH           FACTORS INFLUENCING VASCULAR PLANT ZONATION IN
                 NORTH CAROLINA SALTMARSHES
                                   AD-290 938(K)  $7.60 0603
NORTH           SONAR STUDIES OF THE DEEP SCATTERING LAYER IN
                THE NORTH PACIFIC   PB 162 427(K)  $2.60 0587
NORTH           THE DEVELOPMENT OF RESCUE AND SURVIVAL TECHNIQ
                UES IN THE NORTH AMERICAN ARCTIC
                                    PB 162 410(K) $12.00 [illegible]
NOSE            THE FLORA OF HEALTHY DOGS. I. BACTERIA AND FUN
                GI OF THE NOSE, THROAT, AND LOWER INTESTINE
                                        LF-2(K)  $2.60 [illegible]
NOZZLE          FABRICATION OF PYROLYTIC GRAPHITE ROCKET NOZZL
                E COMPONENTS        PB 162 371(K)  $1.10 [illegible]
NOZZLE          FABRICATION OF PYROLYTIC GRAPHITE ROCKET NOZZL
                E COMPONENTS        PB 162 370(K)  $1.10 [illegible]
NOZZLE          FABRICATION OF PYROLYTIC GRAPHITE ROCKET NOZZL
                E COMPONENTS        PB 162 372(K)  $2.60 [illegible]
NOZZLE          THIRD SYMPOSIUM ON ADVANCED PROPULSION CONCEPT
                S SPONSORED BY UNITED STATES AIR FORCE OFFICE
                OF SCIENTIFIC RESEARCH AND THE GENERAL ELECTRI
                C COMPANY FLIGHT PROPULSION DIVISION CINCINNAT
                I, OHIO OCTOBER 2-4, 1962. PLASMA FLOW IN A MA
                GNETIC ARC NOZZLE  AD-290 [illegible]  $2.60 [illegible]
NOZZLES         HEAT TRANSFER AND PARTICLE TRAJECTORIES IN SOL
                ID-ROCKET NOZZLES  AD-289 [illegible]  [illegible]
NROTC           DEVELOPMENT AND STANDARDIZATION OF FORMS 3 AND
                 4 OF THE NROTC COMPLETE STUDENT SELECTION TES
                T                  AD-290 [illegible]  $1.10 [illegible]
NROTC           EVALUATION OF NROTC AVIATION INDOCTRINATION FI
                ELD TRIPS FOR 1961-1962
                                   AD-290 [illegible]  $1.60 [illegible]
NUCLEAR         A TIME CODE FOR THE CALCULATION OF ELECTROMAGN
                ETIC BLACKOUT FOLLOWING A HIGH ALTITUDE NUCLEA
                R DETONATION       AD-291 141(K)  $8.60 0372
NUCLEAR         ACCURATE NUCLEAR FUEL BURNUP ANALYSES
                                    GEAP-4062(K)  $1.60 0362
NUCLEAR         APPLICATION OF NUCLEAR POWER SUPPLIES TO SPACE
                 SYSTEMS            TID-17306(K)  $8.60 0741
NUCLEAR         CAROLINAS-VIRGINIA NUCLEAR POWER ASSOCIATES, I
                NC., RESEARCH AND DEVELOPMENT PROGRAM QUARTERL
                Y PROGRESS REPORT FOR THE PERIOD APRIL - JUNE
                1962                  CVNA-156(K)  $6.60 0839
NUCLEAR         COMPUTER PROGRAMS FOR OPTIMUM START-UP OF NUCL
                EAR PROPULSION SYSTEMS
                                    TID-16730(K)  $1.10 0712
NUCLEAR         DOSE-TIME-DISTANCE CURVES FOR CLOSE-IN FALLOUT
                 FOR LOW YIELD LAND-SURFACE NUCLEAR DETONATION
                S                   PB 162 516(K)  $1.60 0573
NUCLEAR         EXTRUDED CERAMIC NUCLEAR FUEL DEVELOPMENT PROG
                RAM                ACNP-62550(K)  $4.60 0092
NUCLEAR         FEASIBILITY DETERMINATION OF A NUCLEAR THERMIO
                NIC SPACE POWER PLANT
                                   AD-290 068(K)  $2.60 0031
NUCLEAR         HIGH - ENERGY NUCLEAR PHYSICS RESEARCH PROGRAM
                                   AD-291 14 (K)  $1.60 0374
NUCLEAR         HIGH-ENERGY NUCLEAR REACTIONS OF NIOBIUM WITH
                INCIDENT PROTONS AND HELIUM IONS
                                   UCRL-10461(K)  $2.25 0222
NUCLEAR         INVESTIGATIONS ON THE DIRECT CONVERSION OF NUC
                LEAR FISSION ENERGY TO ELECTRICAL ENERGY IN A
                PLASMA DIODE       AD-290 727(K)  $9.60 0385
NUCLEAR         NUCLEAR SUPERHEAT DEVELOPMENT PROGRAM
                                    GNEC-254(K) $14.00 0386
NUCLEAR         PRODUCTION OF TRITIUM BY CONTAINED NUCLEAR EXP
                LOSIONS IN SALT. I. LABORATORY STUDIES OF ISOT
                OPIC EXCHANGE OF TRITIUM IN THE HYDROGEN-WATER
                 SYSTEM             ORNL-3334(K)   $.50 0617
NUCLEAR         STRIKING EFFECT OF NUCLEAR EXPLOSION
                                   AD-290 824(K) $21.00 0083
NUCLEAR         THE NUCLEAR PROPERTIES OF RHENIUM
                                   AD-291 180(K)  $1.60 0310
NUCLEAR         VARIATIONS IN THE TOTAL ELECTRON CONTENT OF TH
                E IONOSPHERE AFTER THE HIGH ALTITUDE NUCLEAR E
                XPLOSION         NASA N63-10486(K)  $1.10 0142
NUCLEAR         630A MARITIME NUCLEAR STEAM GENERATOR
                                     GEMP-160(K)  $8.10 0349
NULL-ZONE       THE ESTIMATION PROBLEM IN NULL-ZONE RECEPTION
                FEEDBACK SYSTEMS   AD-290 325(K) $11.00 0599
NUMBERS         FUNDAMENTAL SOLUTION TO THE DIFFUSION BOUNDARY
                 LAYER EQUATION FOR NEARLY SEPARATED FLOW OVER
                 SOLID SURFACES AT VERY LARGE PRANDTL NUMBERS
                                   AD-291 031(K)  $2.60 0023
NUMBERS         LOCAL PRESSURE DISTRIBUTION ON A BLUNT DELTA W
                ING FOR ANGLES OF ATTACK UP TO 35-DEGREES AT M
                ACH NUMBERS OF 3.4 AND 4.7
                                 NASA N63-10800(K)   $.75 0516
NUMERICAL       A MAINTENANCE PROGRAM FOR NUMERICAL CONTROL SY
                STEMS ON MACHINE TOOLS
                                    TID-17376(K)  $2.60 0809
NUMERICAL       A PRIORI BOUNDS ON THE SISCRETIZATION ERROR IN
                 THE NUMERICAL SOLUTION OF THE DIRICHLET PROBL
                EM                 AD-290 322(K)  $4.60 0464
NUMERICAL       NON-SIMILAR NUMERICAL METHODS OF SOLUTION FOR
                ELECTRODE BOUNDARY LAYERS IN A CROSSED FIELD A
                CCELERATOR         AD-290 525(K)  $5.60 0185
NYSTAGMUS       MANIPULATION OF AROUSAL AND ITS EFFECTS ON HUM
                AN VESTIBULAR NYSTAGMUS INDUCED BY CALORIC IRR
                IGATION AND ANGULAR ACCELERATIONS
                                   AD-290 348(K)  $1.60 0252
OAK             A SAFETY REVIEW OF THE OAK RIDGE CRITICAL EXPE
                RIMENTS FACILITY   ORNL-TM-349(K)  $5.60 0612
OBJECTS         DRAG OF OBJECTS IN PARTICLE - LADEN AIR FLOW P
                HASE IV. BLUNT BODIES AND COMPRESSIBILITY EFFE
                CTS                AD-291 170(K)  $6.60 0752
OBSERVATORY     TONTO FOREST SEISMOLOGICAL OBSERVATORY
                                   AD-291 140(K)  $3.60 0813
OCEAN           A SAMPLE TEST EXPOSURE TO EXAMINE CORROSION AN
                D FOULING OF EQUIPMENT INSTALLED IN THE DEEP O
                CEAN               AD-291 049(K)  $1.60 0582
OCEANOGRAPHIC   OCEANOGRAPHIC CRUISE TO THE BERING AND CHUKCHI
                 SEAS, SUMMER 1949. PART I SEA FLOOR STUDIES
                                    PB 162 420(K)  $2.60 [illegible]
OCEANOGRAPHIC   OCEANOGRAPHIC AND UNDERWATER ACCOUSTICS RESEAR
                CH                 AD-290 252(K)  $2.60 [illegible]
OCEANOGRAPHIC   OCEANOGRAPHIC CRUISE TO THE BERING AND CHUKCHI
                 SEAS, SUMMER 1949. PART IV. PHYSICAL OCEANOGR
                APHIC STUDIES. VOL. I. DESCRIPTIVE REPORT
                                  PB 162 420-1(K)  $3.60 [illegible]
OCEANOGRAPHIC   OCEANOGRAPHIC CRUISE TO THE BERING AND CHUKCHI
                 SEAS, SUMMER 1949. PART IV. PHYSICAL OCEANOGR
                APHIC STUDIES. VOL. I. DESCRIPTIVE REPORT
                                  PB 162 420-1(K)  $3.60 [illegible]
OCEANOGRAPHIC   OCEANOGRAPHIC CRUISE TO THE BERING AND CHUKCHI
                 SEAS, SUMMER 1949. PART IV PHYSICAL OCEANOGRA
                PHIC STUDIES. VOL. 2. DATA REPORT
                                  PB 162 420-2(K)  [illegible]
OCEANOGRAPHIC   OCEANOGRAPHIC CRUISE TO THE BERING AND CHUKCHI
                 SEAS, SUMMER 1949. PART IV PHYSICAL OCEANOGRA
                PHIC STUDIES. VOL. 2. DATA REPORT
                                  PB 162 420-2(K)  [illegible]
OCEANOGRAPHIC   PROCEEDINGS OF [illegible] OCEANOGRAPHIC
                [illegible] BURBANK, CALIFORNIA, 9 JUNE
                 1962               PB 162 [illegible]  [illegible]
[illegible]     RUBBER ELASTICITY IN HIGHLY CROSSLINKED SYSTEM
```

Figure 6. Sample, CEIR Format for Office of Technical Services



DIABETES RELATED LITERATURE KEYWORD INDEX 1960

CHANGES OF GLYCEMIA IN THE UMBILICAL VEIN FOLLOWING INTRAVENOUS ADMINISTRATION OF GLUCOSE TO MOTHER. • Z K STEMBERA, J HODR • CESK GYNEK V24 P610-6, OCT 59 CZ

MODIFICATION OF THE GLYCEMIA LEVEL, PYRUVIC ACID LEVEL AND THE LEVEL OF INORGANIC PHOSPHORUS BY APPLICATION OF GLUCOSE DURING LABOR WITH CONSIDERATION TO HYPOXIA OF THE FETUS. • J HODR, J HERZMANN, J JANDA • Z GEBURTSH GYNAEK V154 P57-75 1959 GER

EFFECTS OF THE ADMINISTRATION OF SULFONAMIDE BY WAY OF THE EXOCRINE DUCTS ON THE GLYCEMIA AND HISTOLOGICAL STRUCTURE OF THE PANCREAS. • A LOUBATIERES, A SASSINE, M M MARIANI, C FRUTEAU DE LACLOS • C R SOC BIOL PAR V154 P155-8, 1960 FR

EFFECT OF SODIUM ACETOACETATE ON GLYCEMIA. • M TOTH, L BARTA • ACTA MED ACAD SCI HUNG V15 P343-6, 1960 FR

MODIFICATIONS IN GLYCEMIA AND GLUCOSE LOADING CURVE IN ANIMALS WITH CHRONIC LESIONS OF THE SPINAL CORD • G PINNA, M B DECHERCHI • BOLL SOC ITAL BIOL SPER V35 P1885-8, 31 DEC 59 IT

THE GLYCEMIC CYCLE COMPARED WITH THE INDUCED HYPERGLYCEMIA TEST AND THE FASTING GLYCEMIA, ITS IMPORTANCE IN DIABETICS, EVEN IN THOSE APPARENTLY IN EQUILIBRIUM, SIMPLIFIED PERFORMANCE OF TEST USING AUTO-MICRO-SAMPLINGS. • C PEREZ • TUNISIE MED V38 P199-209, MAR 60 FR

EFFECT OF OVERSTIMULATION OF THE CNS ON GLYCEMIA IN RATS IN VARIOUS CONDITIONS. • M SUBOVA • CESK FYSIOL V8 P556, NOV 59 CZ

THE GLYCEMIC CYCLE COMPARED WITH THE INDUCED HYPERGLYCEMIA TEST AND THE FASTING GLYCEMIA, ITS IMPORTANCE IN DIABETICS EVEN IN THOSE APPARENTLY IN EQUILIBRIUM, SIMPLIFIED PERFORMANCE OF TEST USING AUTO-MICRO-SAMPLINGS. • C PEREZ • TUNISIE MED V38 P199-209, MAR 60 FR

EFFECT OF INSULIN ON GLYCEMIA STUDIES BY MEANS OF TEMPORARY AND PERMANENT METHODS OF LIGATION OF THE V. PORTAE AND V. HEPATICUM IN RATS. • R KOREC • CESK FYSIOL V9 P28, JAN 60 CZ

GLYCEMIC

THE AMINOACIDEMIC AND GLYCEMIC RESPONSE IN ULCER PATIENTS AFTER INTRAVENOUS LOAD OF AMINO ACIDS. • I GIORGIO, V OLIVA • BOLL SOC ITAL BIOL SPER V35 P1864-8, 15 SEPT 59 IT

EFFECTS OF CHLORPROMAZINE ON CERTAIN GLYCEMIC TESTS IN CHILDREN. • V TISCHLER, J JACINA, B HRUBA, O PAVKOVCEKOVA CESK PEDIAT V14 P677-89, AUG 59 CZ

THE GLYCEMIC CYCLE COMPARED WITH THE INDUCED HYPERGLYCEMIA TEST AND THE FASTING GLYCEMIA, ITS IMPORTANCE IN DIABETICS, EVEN IN THOSE APPARENTLY IN EQUILIBRIUM, SIMPLIFIED PERFORMANCE OF TEST USING AUTO-MICRO-SAMPLINGS. • C PEREZ • TUNISIE MED V38 P199-209, MAR 60 FR

NEURAL REGULATION OF INDUCED GLYCEMIC REACTION. • E GUTMANN, B JAKOUBEK • CESK FYSIOL V8 P404-5, SEPT 59 CZ

GLYCEMIC CURVE

CHANGES OF GLYCEMIC CURVE FOLLOWING THE ADMINISTRATION OF GALACTOSE IN HEAD INJURIES. • I HAVLIN • CESK FYSIOL V8 P317, JULY 59 CZ

GLYCERIC ACID D

PHOSPHORYLATION OF D-GLYCERIC ACID TO 2-PHOSPHO-D-GLYCERIC ACID WITH GLYCERATE KINASE IN THE LIVER. I. ON THE BIOCHEMISTRY OF FRUCTOSE METABOLISM. II. • W LAMPRECHT, T DIAMANTSTEIN, F HEINZ, P BALDE • HOPPE SEYLER Z PHYSIOL CHEM V316 P97-112, 30 SEPT 59 GER

GLYCERIDE

INFLUENCE OF INSULIN ON THE INCORPORATION OF 2-14 C-SODIUM PYRUVATE INTO GLYCERIDE GLYCEROL IN DIABETIC AND NORMAL BABOONS. • N SAVAGE, J GILLMAN, C GILBERT • NATURE LOND V185 P168-9, 16 JAN 60

GLYCERIDE GLYCEROL

METABOLIC ROLE OF GLUCOSE. A SOURCE OF GLYCERIDE-GLYCEROL IN CONTROLLING THE RELEASE OF FATTY ACIDS BY ADIPOSE TISSUE. • F C WOOD JR., B LEBOEUF, G F CAHILL JR. • DIABETES V9 P261-3, JULY-AUG 60

GLYCEROL

EFFECT OF EPINEPHRINE ON GLUCOSE UPTAKE AND GLYCEROL RELEASE BY ADIPOSE TISSUE IN VITRO. • B LEBOEUF, B FLINN, G F CAHILL JR. • PROC SOC EXP BIOL MED V102 P527-9, OCT-DEC 59

INFLUENCE OF INSULIN ON THE INCORPORATION OF 2-14 C-SODIUM PYRUVATE INTO GLYCERIDE GLYCEROL IN DIABETIC AND NORMAL BABOONS. • N SAVAGE, J GILLMAN, C GILBERT • NATURE LOND V185 P168-9, 16 JAN 60

UNIMPAIRED SYNTHESIS OF FATTY ACIDS AND ALTERED SYNTHESIS OF GLYCEROL OF TRIGLYCERIDES IN DIABETIC BABOONS P. URSINUS. • N SAVAGE, J GILLMAN, C GILBERT • S AFR J MED SCI V25 P19-32, APR 60

GLYCIDE

MATERNAL GLYCIDE NORMAL ASSIMILATION, TOMATO BABY, PRECEDENTS OF MACROSOMIA AND FETAL MORTALITY. • B SALVADORI, G CAGNAZZO, A DELEONARDIS • MINERVA PEDIAT V12 P117, 11 FEB 60 IT

GLYCINE

AN INSULIN ASSAY BASED ON THE INCORPORATION OF LABELLED GLYCINE INTO PROTEIN OF ISOLATED RAT DIAPHRAGM. • K L MANCHESTER, P J RANDLE, F G YOUNG • J ENDOCR V19 P259-62, DEC 59

MAINTENANCE OF CARBOHYDRATE STORES DURING STRESS OF COLD AND FATIGUE IN RATS PREFED DIETS CONTAINING ADDED GLYCINE. • W R TODD, M ALLEN • USAF ARCTIC AEROMED LAB TECHN REP V57-36 P1-16, JUNE 60

GLYCINE C14

Figure 7. Sample Page, Diabetes Index

essentially the original Luhn format, and it should be noted in this connection that while Luhn recognized that the origin of the KWIC principle lay in the making of concordances, he claimed in particular the use of machines to achieve speed, completeness, and accuracy, and a novel format. 1/

The most common variant to the center position for the indexing window (or keyword position) is at the left or the beginning of the line. Netherwood's selected bibliography of logical machine design, which is probably the first of the modern permuted title indexes to appear in the open literature, used the left-most positions for the index entry word in each title listing. Slant marks were also printed to show the breaks in the normal order of the title (Netherwood, 1958 [437]). A proposed subscription service, advertized in 1958 but never actually brought into operation, would also have used the left-hand position. 2/

In these left position examples, the keyword-in-context principle is kept only partially intact since the word in the index position is directly adjacent to its most specific right-hand context, not to its left-hand. In variations such as developed at Stanford Research Institute, however, the index word is extracted from its context and printed separately in the left-hand margin, with the title in its normal order printed to the right. This type of variation has been called "KWOC", for keyword-out-of-context, and is illustrated in Figure 6, which shows the format developed by C.E.I.R., Inc. for the OTS index to U.S. Government Research Reports.

Table 1 lists a number of KWIC index projects for which computer programs are or might be made available to interested additional users. Computer programs have been written specifically for the IBM 650, 704, 1620, 709, 7090, and 7094 data processing systems, the G.E. 225 computer, the Deuce Computer in England, the UNIVAC 1103 and 1107 systems, and the Japanese computer JEIPAC, among others. In addition, some permuted title indexes are produced manually, or with the use of simple business office machine equipment. For example, an index to the AIBS Bulletin for 1951-1961 has been so produced by the American Institute of Biological Sciences. 3/

1/ Private communication, excerpt of letter from H. P. Luhn to C. L. Bernier, December 27, 1960: "With respect to the origin of the KWIC Index, you are, of course, right that it is a form of concordance, as stated in my original paper. Furthermore, keyword indexing has been practiced in various forms as far back as a hundred years ago. All of these methods were, however, dependent on manual effort. I would say that the significance of the present KWIC Index is based on the fact that it is produced automatically by machine, affording speed of compilation, accuracy and completeness. As far as the particular format of the Index is concerned, this is novel to my knowledge, in accordance with information I have been able to ascertain from others."

2/ "PILOT--a permutation index to this month's literature", see p. 8 and Figure 1. A left-most window full-title format was developed at Stanford University in cooperation with the IBM San Jose Laboratories. It has been applied by the Computation Center to the titles of computer programs for the benefit of users of the Program Library   Computation Center, Stanford University, "The KWIC Index", 1963. See also Marckworth, 1961 [393].

3/ National Science Foundation's CR&D Report No. 11, [430], p.10; Janaske, 1962 [299]; Shilling, 1963 [550] and [551].

Table 1. KWIC Type Indexes and Programs

Issuing Organization and/or Investigator	Name of Index or Program	When Issued	Format	Computer	References and Remarks
Service Bureau Corporation - H. P. Luhn	"Bibliography and Auto-Index, Literature on Information Retrieval and Machine Translation"	First edition Sept. 1958 Second edition June 1959	2-column, 60-character single line title, center window	IBM 709	Basic Luhn KWIC
Chemical Abstracts Service	Chemical Titles	Semi-monthly	Standard Luhn IBM	1401	
Chemical Abstracts Service	Chemical Biological Activities	Bi-weekly-1st issue Sept. 1962	Single column Center window, 120-character line, upper and lower case, 120-character 1403 printer	1401	
Biological Abstracts	B. A. S. I. C.	Semi-monthly	Standard Luhn IBM	1440	Modified Luhn program: shading is used as an aid in scanning.
Biological Abstracts	Biochemical Title Index	Monthly	Luhn, Chem. Titles Formats	1440	
Bell Telephone Laboratories	-Index to the Literature of Magnetism -BTL talks and papers	Annually Annually	Single column, 120 character line, center window	7090	BE-PIP Program available through the SHARE organization
All-Union Inst. for Scientific and Technical Information			"... an index of the 'Chemical Titles' type."		Mikhailov, 1962 [418]

Table 1. (cont.)

Issuing Organization and/or Investigator	Name of Index or Program	When Issued	Format	Computer	References and Remarks
American Bar Foundation, Bobbs Merrill	Index to Current State Legislation	Initial issue, 1963			Eldridge and Dennis, 1962 [183]
American Diabetes Association	Diabetes-related Literature Index	First of proposed series, covering literature for 1960, issued 1963	2-column, left window, KWOC, full citation for each entry.	GE-225, Western Reserve program	
American Meteorological Society	Meteorological and Geoastrophysical Titles	April 1961, Oct. 1961, Jan. 1962 and following	Standard Luhn IBM	704	Includes a Systematic UDC-Subject Heading Index as well as modified KWIC.
Armour Research Foundation	Key words in context (reports received in document library)		Two column, 60-character line, center window	1103, 1107	
ASTIA (Defense Documentation Center)	Keywords-in-context title index. A list of titles for ASTIA documents not previously announced.	Irregularly No. 1, Oct. 1962 No. 2, Feb. 1963		IBM	
English Electric Company			KWIC-type	Deuce	See Black, 1962 [65]; Dowell and Marshall, 1962 [159].

Table 1. (cont.)

Issuing Organization and/or Investigator	Name of Index or Program	When Issued	Format	Computer	References and Remarks
General Electric Computer Dept. Phoenix	General Bibliography on Information Storage and Retrieval	- - -	Single column, center window	GE-225	
Gmelin Institute	Information Journal for Atomic Energy				See Koelewijn, 1962 [330].
Japan Information Center of Science and Technology				JEIPAC	"The JEIPAC, a transistorized information processing machine... has also been programmed for automatic indexing designed after the IBM KWIC indexing system." CR & D No. 11, [430], p. 120-121.
Lockheed Missiles and Space Division	KWIC Index of Reports		Modified Bell Labs.	1401/ 7090	See Carroll and Summit, 1962 [102].
Mimosa Frenk Foundation for Applied Neurochemistry	KWIC Index to Neurochemistry	August 1961	Standard Luhn IBM	IBM	
M. I. T.	KWIC Index to The Science Abstracts of China	1st Edition, December 1960	Standard Luhn IBM	704	

Table 1. (cont.)

Issuing Organization and/or Investigator	Name of Index or Program	When Issued	Format	Computer	References and Remarks
National Bureau of Standards	A Bibliography of Foreign Developments in Machine Translation and Information Processing	July 1963	Single column, 120-character line, center window	7090	Byproduct input from Flexowriter tape, citation data including upper and lower case, paper tape to punched card conversion. Walkowicz, 1963 [629].
National Bureau of Standards, W. W. Youden	-Index to the Communications of the ACM -Index to The Journal of the ACM		Single column, 120-character line, center window	7090	Youden, 1963 [659] and [660].
Radio Corporation of America	Significant Words Indexed From Title			RCA 301	Unpublished report by D. Climenson and M. Bechman
Stanford Univ. IBM San Jose Labs.	Dissertations in Physics	1961	Keyword-out-of-context, left window	IBM	Marckworth, 1961 [393].
Union Carbide Oak Ridge National Laboratory Libraries	Key Word Index Laboratory Reports Received Semi-annual Index January-June 1963	1st issue 1963, monthly thereafter	Bell Labs. System		
U.S. Atomic Energy Commission, Division of Technical Information	Index to Conferences Abstracted in Nuclear Science Abstracts	December 1963	Bell Labs. System		

Table 1. (cont.)

Issuing Organization and/or Investigator	Name of Index or Program	When Issued	Format	Computer	References and Remarks
University of California Lawrence Radiation Laboratories	Key-word-in-title (KWIT) index for reports	Various issues	Single column, 120-character line, center window	1401/7090	Records can be machine searched with and, or and not logic
University of California, Lawrence Radiation Laboratories	Unclassified Reports Titles List	Biweekly	Single column, 120-character line, center window	1401	By-product preparation from Flexowriter of library cards. Turner and Kennedy, 1961 [614].
University of Kansas	Kansas Slavic Index	Initial issue July 1963	60-character Modified Chemical Titles	1401	Farley, 1963 [192].
University of Kansas, University of Oklahoma	(Space Law collection)			1401	"Current research and development..." No. 11, p. 44 & 171.
Western Periodicals Company	Permuted Indexes to Scientific Symposia	As available	Standard Luhn IBM		Advertised regularly in various periodicals, e.g., Special Libraries

In addition to the regularly issued KWIC indexes by Biological Abstracts, Chemical Abstracts Service, the American Meteorological Society and others, a large number of special field, one time, or limited collection coverage indexes of this type have been and are being produced both in the United States and in other countries. Well-known examples include the programs developed at the Lawrence Radiation Laboratories, University of California, which simultaneously produce catalog, cross-reference and subject authority cards, [1] and the programs developed at the Bell Telephone Laboratories from 1959 onward (Kennedy, 1962 [310]).

Other KWIC indexing efforts cover a wide variety of subject matter. In the field of law, applications of KWIC type indexing include work on the legislation of the 50 states, a joint project of the American Bar Foundation and the Bobbs-Merrill Company (Eldridge and Dennis, 1962 [183], 1963 [182]), the ninth annual edition of the Index to Legal Theses and Research Projects, July 1962, (Eldridge and Dennis, 1963 [182]); and a cooperative program between the libraries of the Universities of Kansas and Oklahoma to prepare an index to the latter's "Space Law" collection. [2] In 1960, the KWIC Index to the Science Abstracts of China was prepared for an AAAS Symposium, (Henderson, 1961 [263]; Farley, 1963 [192]). At the University of Kansas Library also, the Kansas Slavic Index is being produced, with coverage of 3,000 articles from more than 200 Slavic journals. [3] In the computer technology field, Youden (1963 [659] and [660]) has compiled KWIC type indexes to both the Journal of the ACM and the Communications of the ACM and the Western Periodicals Company offers KWIC indexes to the proceedings of the Joint Computer Conferences as well as to the proceedings of other conferences and symposia including those in fields of electronics, aerospace and quality control. [4] A special-purpose application is in the use of a KWIC-index in lieu of cross-references in a revised edition of Current Medical Terminology. [5]

Examples of KWIC indexing projects abroad include work at the Japanese Information Center of Science and Technology, Tokyo, [6] an index "of the 'Chemical Titles' type" at the All-Union Institute for Scientific and Technical Information (VINITI) U.S.S.R., [7] an information journal for the atomic energy field being prepared at the Gmelin Institute, (Koelwijn, 1962 [330]), and work in Great Britain both at the English Electric Company [8] and the IBM British Laboratories (Black, 1962 [65]).

---

[1] Nation Science Foundation's CR&D Report, No. 11, [430], p. 42.

[2] Ibid, pp. 44 and 171.

[3] Ibid, p. 43; University of Kansas, 1963 [307].

[4] See advertisements in journals such as American Documentation.

[5] Gordon and Slowinski, 1963 [236], p. 55.

[6] National Science Foundation's CR&D Report, No. 11, [430], p. 120.

[7] Mikhailov, 1962 [418], p. 50.

[8] Dowell and Marshall, 1962 [159], p. 323; Black, 1962 [65], p. 316.

Trans-Canada Air Lines [1] is using a KWIC System, and at the EURATOM ISPRA laboratories a KWIC type program has been developed with up to 600-character context and a left-most indexing position. [2]

3.1.2 Advantages, Disadvantages and Operational Problems of KWIC Indexing

Luhn's original acronym, KWIC, is peculiarly apt for permuted title word indexing. As both proponents and critics have noted, the resulting product may be relatively crude in terms of indexing quality, but it is quick. The speed achievable both by elimination of human intellectual effort and by use of machine (especially computer) processing is indeed the major single advantage of this type of automatic indexing. Closely related, however, are the advantages of currency of announcement and the availability of these indexes for individual use.

Some typical claims with respect to speed and currency are as follows:

> "The permuted index was invented as a means of adequately controlling (essentially, of indexing) the literature without further intellectual effort, and thus eliminating indexing delays." [3]
>
> "The great merit of this particular method... is that it enables information concerning new articles to be made available very much more quickly than if there were the inevitable delays of human abstracting and indexing." [4]
>
> "In spite of the disadvantages which are pointed out, perhaps the greatest advantage is the timeliness and the speed with which permuted-title indexes can be prepared." [5]

Specific examples of high speed are given by Biological Abstracts, where one hour's computer time suffices to prepare and arrange entries for over 150,000 items. [6] Kennedy reports for the Bell Laboratories System that:

> "Editorial scanning is very fast; only several lines of print must be read for each report and the required text markings are trivially few. Keypunching, the largest single task, takes about two minutes per report... Main-frame time ...was 12 minutes for 1703 reports." [7]

---

[1] Simons,1963 [556], p. 34.

[2] Meyer-Uhlenried and Lustig, 1963 [417], p.229.

[3] Tukey, 1962 [611], p.13.

[4] Cleverdon, 1961 [125], p.108.

[5] Janaske, 1962 [299], p.3.

[6] See Biological Abstracts, 36:24, p. xii.

[7] Kennedy, 1961 [311], p.123.

Skaggs and Spangler claim:

> "The most obvious advantage of permuted indexing by computer is speed. In a test of one permuted indexing system, input of 3,000 punched cards containing titles and running text produced a permuted significant word index of 12,190 index entry lines, with approximately 85 minutes of computer time required for the permuting and sort operations. The output was printed at some 500 lines per minute..." 1/

In many cases, greater speed and timeliness are achieved at significantly lower cost. This is particularly true if the preparation of the input -- title, author, item identification and other descriptive cataloging information--serves multiple purposes from a single keystroking operation. Thus, the MATICO System provides from a single input (1) KWIC indexes as required, (2) selective dissemination notices to potential users of new acquisitions, (3) records on magnetic tape for the information retrieval file, and (4) book catalogs covering specialized areas of the collection, all at a net savings over previous methods of $0.39 for each title processed. 2/

Another advantage which is typically claimed for KWIC indexes is the use of the author's own terminology. The display of different words as they have been used in title context with any word looked up introduces "suggestiveness" so that different meanings and different browsing clues are shown. Kennedy makes the following typical points:

> "The use of the author's own terms--the alive currency of new ideas--rather than the considered reshapings to the indexing system may often be of advantage. The automatic generation as index entries of all the separate words in multi-term concepts is definitely so. Access is direct, under any one of the component terms, in the unrestricted manner of uniterm indexing. And context minimizes false drops; the author has supplied the term coordination." 3/

Others, however, consider some of these same factors to be definite disadvantages.

In general, even among enthusiasts of KWIC, there is more agreement as to the values of the technique as a device for current awareness scanning and as a dissemination index than for its use for more extensive searching. It was, in fact, primarily as a dissemination index that Luhn first proposed the KWIC technique. He pointed out that such indexes could be prepared with minimum effort and be ready for dissemination in the shortest possible time, justifying publication by inexpensive printing means. He also noted the following additional advantages:

1/ Skaggs and Spangler, 1963 [557], p. 30.

2/ Carroll and Summit, 1962 [102], p. 4.

3/ Kennedy, 1962 [310], p. 184.

"1. Because of the mechanical method of preparation, more information may be displayed than would have been practicable by conventional means.

"2. Keywords-in-context permit the cross-correlation of subjects to an extent not realizable by conventional procedures." [1]

The most common type of complaint against the KWIC indexing method is, as we have noted earlier, identical with that which is applied to word indexing in general--the lack of terminological control. Where the indexing terms are restricted to those used by the author himself, in his title or even full text, there arise many serious problems of synonyms, near-synonyms, homographs, neologisms, and eponyms. The effects of machine inability to resolve these problems are redundancy, scatter of references throughout the index, "haphazard groupings", [2] and retrieval losses because the user is forced to guess at the terminology the author actually used. [3] These problems are severely aggravated when only the title is used as the basis for index-word extraction.

Thus, a first and major question in attempting to appraise the effectiveness of KWIC-indexing techniques is that of the adequacy of titles alone as the source of subject content clues. Spurred on at least in part by the existence of KWIC-type indexes, several investigators have studied this question, with somewhat different results. Williams has explored for some years the possibilities of developing systematic procedures for title elaboration, especially making explicit information that is implied. Her conclusions are that indexing by title and direct elaboration of the title would produce index information equivalent to that found in Chemical Abstracts for about 50 percent of the documents studied, but that other procedures would be required for the remainder. [4]

Specific studies of title adequacy for a particular journal or field have been undertaken by both the American Institute of Physics and the Biological Sciences Communications Project. In the A. I. P. experiments, graduate physics students were asked to locate from limited clues certain specific articles appearing in The Physical Review, and search times were checked for their use of permuted title and other indexes. Another group of students compared the subject index entries in Physics Abstracts and Chemical Abstracts with the words in the titles of 25 papers from The Physical Review. In the case of Physics Abstracts, 69 percent of the entries for these papers were found in the words of the title and 63 percent of the titles contained all of the information supplied by the set of index entries. In the case of Chemical Abstracts, the corresponding percentages were 47 and 23. [5] These latter findings, for the chemical index, are closely corroborated

---

[1] Luhn, 1959, [381], p. 295.

[2] Olney, 1963, [458], p. 44.

[3] See, for example, Dowell and Marshall, 1962, [159], p. 324: "This problem of 'conceptual scatter' becomes a nightmare when highly idiosyncratic author language is used as a basis for subject indexing."

[4] Williams, 1961 [643], pp. 361-363.

[5] Maizell, 1960 [392], p. 126.

Bernier and Crane who report that for the non-organic chemistry items covered by Chemical Abstracts, 34 percent of the entries can be derived from the titles. [1]

With respect to the Biological Sciences Communications Project studies, Shilling reports as follows:

> "Titles of scientific articles are being utilized at present in a great many ways under the general assumption that there is a positive correlation between the title and the content of the article. A study was undertaken to analyze the accuracy of titles in describing the content of biomedical articles. It was conducted in two parts. In part one, a group of scientists were asked to predict the content of selected scientific articles, in their area of interest, from the title, the author's name, and the name of the journal in which it appeared. The results of the first phase of the study on the first trial journal were so diverse as to make analysis impossible, and this part of the study was not pursued further. From this small segment of the study it appears that scientists are deluding themselves when they search by title only and then decide what they wish to read.
>
> "In the other half of this experiment, the article without title, author's name, or journal name was sent to 20 scientists, selected as experts in the scientific field of the article, who were asked to write a meaningful title. Fifty articles were used, five from each of ten selected biomedical journals. From this part of the study it is apparent that if the article is in a field which is relatively well standardized and has an accepted vocabulary, it is possible for a group of titlists to agree remarkably well on an appropriate title. However, if the article is loosely organized, contains more than one subject, or is in a specialty in which there is no standard vocabulary, then titling scientists fail to agree to a rather alarming extent."[2]

Other studies involving the question of usefulness of titles alone for indexing purposes include those of Doyle, Lane, Montgomery and Swanson, O'Connor, Ruhl, Swanson, and White and Walsh, among others. Doyle checked the retrieval loss likely to result from the synonymity-scatter problem for a permuted title index compiled in 1958 to the internal reports of the System Development Corporation. He found, for example, that for 12 direct references to McGuire Air Force Base, there were one to "New York Air Defense Sector", two to "New York Sector", ten to "NYADS" and five to "N. Y. Sector". [3]

---

[1] Bernier and Crane, 1962 [56], p. 120.

[2] Shilling, 1963 [551], pp. 205-206.

[3] Doyle, 1961 [166], p. 11.

Ruhl (1963 [506]) found that between 50 and 90 percent of author-prepared titles (the variation depending on subject field and other circumstances), did fully reflect the index terms assigned to these documents by human indexers. Lane and White and Walsh have also made studies directly related to the question of KWIC index effectiveness. The latter two investigators report only 52 percent retrieval effectiveness for a permuted title index to the Abstracts of Computer Literature, 1962, which they attribute to the changing terminology in the still new field of computer technology. [1] Lane made counts of titles that would be "acceptable" and those that would not for a KWIC index for 50 titles drawn from each of 10 published indexes. He concluded that, if there were judicious pre-editing, technical articles in the technical subject indexes could be quite adequately covered, and papers in the fields of law, business, and the humanities somewhat less satisfactorily so, but that for the material indexed in the Reader's Guide to Periodical Literature, the KWIC technique would fail 58 percent of the time. [2]

Montgomery and Swanson have studied, as has O'Connor is even more detail, the adequacy of "machine-like indexing by people". Montgomery and Swanson took as their test corpus the September 1960 issue of Index Medicus and found that for 4,770 items, 85.8 percent contained either the word itself or a synonym for the subject heading assigned, slightly over 11 percent did not, and in the remaining cases the investigators could not clearly decide. They concluded, therefore, that: "Most of the articles studied could have been indexed by machine on the basis of machine 'inspection' of article titles alone." [3] O'Connor, however, typically reports that of a random sample of 50 papers manually indexed under the term "Toxicity", five had titles which contained the word "toxic" or the word "toxicity" and 34 had titles which were not even indirectly connected with the term. ([443], [444], [445], [447] and [448]). With respect to the Montgomery-Swanson conclusions as such, Carlson raises the further critical questions of over-assignment and false drops and suggests that: "a simple machine processing of titles would give us way too much or practically nothing." [4]

Research activities at the American Bar Foundation have included checking of KWIC type indexing of several thousand legal articles with the subject headings assigned under the "Index to Legal Periodicals" system (Kraft, 1962 [333]). It is reported that:

---

[1] White and Walsh, 1963 [639], p. 346.

[2] Lane, 1964 [345], p. 46.

[3] Montgomery and Swanson, 1962 [421], p. 359. In another study (1962 [534], p. 468), Swanson reports findings for several thousand entries in classified bibliographies where approximately 90 percent of the sampled items contained title words that were identical, or similar in meaning, to the subject headings under which they were indexed. He notes, however, that similar results could have been produced by machine processing with the significant proviso that the machine have available an adequate synonym dictionary or thesaurus.

[4] G. Carlson, 1963 [100], pp. 328-329.

"Interpretation of data revealed, among other things, that 64.4 percent of the title entries contained as keywords one or more of the ILP subject heading words under which they were indexed, and 25.1 percent contained logical equivalents. The remaining 10.5 percent of the title entries had non-descriptive titles." [1]

The difficulties with titles as sources of the indexing information stem from at least three distinct types of determining factors: (1) the language habits, background, interests, and idiosyncracies of the author; (2) the interests, familiarity with the subject matter, language habits, imagination, and idiosyncracies of the user, and (3) factors largely extrinsic to either the particular author or the particular user. In the first case, we find especially the problem of the witty, punning, deliberately non-informative title, the so-called "pathological title". Janske gives the provocative example, in the literature of information selection and retrieval itself, of "The Golden Retriever". [2] Even in the non-pathological case, however, there is the serious question of whether the author himself is likely to be a good indexer. [3]

On the user side, the normal critical problems of "bringing the vocabulary of indexer and searcher into coincidence" (Bernier, 1953 [55]) are aggravated by the facts that the user of KWIC must anticipate the terminology used by a large number of different "indexers" (i.e., the authors), that title words spelled the same but with quite different meanings in different special applications are grouped together in the same place in the index, and that the same concepts may be expressed in quite different phraseology depending on the author's, rather than the user's, field of specialization. To these aggravating circumstances there must be added in turn the psychological acceptability to the individual user of the scatter and redundancy, to say nothing of the format and legibility, of a particular published index.

Such factors affecting the particular user will of course vary with the nature and purpost of his search. Kennedy points out, for example, that the location of a document from only a single clue, a single title word, is particularly easy with a permuted title index and he emphasizes that the "index purpose, use, size, statement and array are other factors of considerable moment in judging the value of title indexes". [4]

---

[1] National Science Foundation's CR&D Report No. 11, [430], p. 62.

[2] Janaske, 1962 [299], p. 4.

[3] See, for example, a report on a conference on better indexes for technical literature, ASLIB Proceedings, 13:4, April 1961, with a number of statements on the author as a poor indexer. See also Crane and Bernier, 1958 [144], p. 515: "Not even authors are qualified to index their own work unless they are equipped for the task by training and experience".

[4] Kennedy, 1961 [311], p. 125.

A major question in the area of user acceptability, however, is that of the adequacy of title alone to tell the searcher whether or not a specific document is relevant to his query or interest. A number of investigators, both documentalists and user-scientists, suggest that this is rarely the case. [1] In fact, for many users, titles alone provide only a negative searching device--in an announcement bulletin or abstract journal the user's scanning of titles merely tells him whether or not he should read the abstract and then perhaps go on to the paper itself.

It is for reasons of this type, in all probability, that Montgomery and Swanson found less effectiveness of titles on relevance-judgment tests than might be suggested by their more optimistic findings as to the success of machine procedures for replicating human subject heading assignments. Whereas they have claimed that about 90 percent of test items could have been as successfully indexed by machine as by manual procedures, (Montgomery and Swanson, 1962 [421]; Swanson, 1962 [584]), they have also reported that: "Comparison of title relevance judgment with judgment based on full text examination indicates that titles are only about one-third effective (i. e., two-thirds of the relevant articles would be judged irrelevant) as the basis for estimating the relevance of the article to a given question". [2] They go on to suggest, therefore, that "... indexing should be based on more than titles and... a bibliographic citation system should present to the requester something more than titles." [3] Similarly, Jahoda reports in an analysis of 281 actual search requests at Esso Research and Engineering that only two-thirds could have been answered with a shallow index based on titles and major section headings of the documents and that answering the remainder of the requests would have required an index of considerable depth. [4]

The obvious factors affecting the utility of titles as the source of indexing-searching clues include, first, the limitation of most titles to the principal subject matter, the main topic or topics of the document. The display of title context does to some extent provide for modifications of the topic to the special aspects treated, but it is of course obvious that a title cannot possibly provide clues to subject content not implied in the words of that title. In many cases, the potential user wants information contained in the paper, or even in its appendices, which was not the principal concern of the author and may not even have been considered significant by him. The claim that the author, who knows his own subject best, has already indexed his paper best by his choice of words and emphasis in text, and especially in his title, is pertinent only to that main subject to which he addresses himself, not to the other potentially useful information which he may also disclose.

---

[1] See, for example, Atherton and Yovich, reporting on evaluations by physicists of experimental citation indexing, 1962 [26], p. 22: "The reliance on titles of papers for retrieval purposes was not sufficient"; Levery, 1963 [359], p. 235. "Titles are usually insufficient to furnish a correct index to the text"; Hocken, 1962 [274], p. 93: "The titles were not explicit enough"; Crane and Bernier, 1959 [145], p. 1053: "Lists of titles can be prepared rapidly, but they are inadequately useful in selecting articles of interest, and they provide little or no directly usable information"; Dowell and Marshall, 1962, [159], p. 324: "Frequently titles either lack sufficient detail or are in fact misleading"; Connolly, 1963 [136], p. 35: "Most titles are inadequate as descriptions of the contents of papers."

[2] Montgomery and Swanson, 1962 [421], p. 364.

[3] Ibid, p. 366.

[4] Jahoda, 1962 [298], p. 75.

Other extrinsic factors affecting title adequacy and hence the effectiveness of title-indexes are the size and the relative homogeneity or heterogeneity of the collection or set of documents so indexed, the breadth or narrowness of the subject field or fields covered, the time period covered and whether for one or many fields. Whether or not material in more than one language is included is a special factor. These various factors interact in various ways, usually with disadvantageous effects when even the most "nondescript" human indexer (that is, one who accepts only words from the text itself) is replaced by "a keypunch operator whose job it is to convert the keywords into machine-readable form, and a machine whose job it is to assimilate machine-readable text and print out its permutations with each significant word serving as an access point." 1/

The difficulties of subject scatter, synonymy, homography, redundancy, and the like, however, will also occur in human indexing that relies heavily on title only, which is perhaps more frequently the case than is generally recognized, 2/ just as much as for machine-generated indexes involving the permutations of keywords in titles. Such disadvantages must therefore be balanced not only against the advantages of speed, timeliness, having an index announcement tool personally available at low cost, and the like, but also against the probability of obtaining as useful a tool within the limits of available human indexing resources and justifiable costs. Cleverdon, for example, comments as follows:

> "There are those who would say that this [KWIC] can in no way be called indexing, and that the value of such indexing must be very much lower than that done by intelligent trained human beings. This is a comfortable thought, but such small evidence as is at present available makes it appear doubtful as to whether it is entirely true. This is not to say that a human being cannot do a better job, but it certainly appears likely that the cost of employing a human being to do it is of doubtful economic value." 3/

1/ Herner, 1962 [266], p. 4.

2/ See, for example, Moss, 1962 [425], p. 39: "I am convinced that a great many of the UDC and other numbers which are provided on millions of cards in technical libraries up and down the country, and which look so erudite, are, in fact, no more than cards transliterating titles, with occasionally similar transliteration of a few randomly chosen words from the abstracts as well. . . We are, in effect, already largely using title indexing and complicating it unnecessarily by magic numbers." See also Crane and Bernier, 1958 [144], p. 514: "Some indexes to periodicals, particularly word indexes, are merely indexes of titles of papers or of abstracts."

3/ Cleverdon, 1961 [125], pp.107-108.

It is also of interest to note, moreover, that the very existence of machine-generated permuted title indexes should greatly increase the likelihood that authors will use better and more useful titles. 1/ At a seminar on word and vocabulary byproducts of permuted title indexing held at Biological Abstracts headquarters on October 8, 1962, Rigby of Meteorological and Geoastrophysical Abstracts reported informally that as of that time there was already discernible improvement in titles covered by their KWIC index. In the same year (1962), Tukey similarly stated that: "Chemical Titles has been heavily enough used to affect the construction of titles of papers on chemical subjects." 2/ Instructions to authors of the previously mentioned "Short Papers" 3/ for the A. D. I. 1963 Annual Meeting specified that at least six significant words should be included in their titles and nearly all authors did in fact comply. Two of the "Short Papers" are specifically directed to the topic of improvements that authors can make in writing their titles (Brandenberg, 1963 [80]; Kennedy, 1963 [312]).

Instructions of this type can be effectively used for situations where all authors are under the same administrative control, as in the internal reports prepared in a single organization. This type of situation, incidentally, is one for which KWIC proponents are often most enthusiastic (Kennedy, 1962 [310]; Black, 1962 [65]; Linder, 1960 [362]). Finally, there is considerable promise that pressures brought to bear by journal editors of the publications of professional societies, notably the American Institute of Chemical Engineers and other cooperating member societies of the Engineers Joint Council, will result in improved adequacy of titles and thereby increased effectiveness of title word indexes.

Certain other disadvantages of KWIC indexing techniques, however, relate specifically to operational problems and requirements in the machine production of these indexes. There is, first, the problem of the amount of context that is usually displayed--that is, the question of line length--and the related problems of title truncation and wrap-around. As Kennedy notes: "Progressive shifting of the title to bring a given word to the indexing column frequently causes portions of the title to exceed the line space available, first at the right margin, then the left, or even both simultaneously." 4/ A case in point is the perhaps apocryphal "EROTIC TENDENCIES AMONG TRAPPIST MONKS" where "ATHEROSCL" had been dropped off at the left.

For multi-column KWIC indexes, in particular, where the line length is typically 58-60 characters, "much of the relevance is lost because the reader sees the wrong slice of the title". 5/ The Bell Laboratories KWIC index, 6/ Chemical-Biological Activities, 7/

1/ See for example, Black, 1962 [65], p. 317; Youden, 1963 [658], p. 332.

2/ Tukey, 1962 [611], pp. 9-10.

3/ Luhn, 1963 [376] and [377].

4/ Kennedy, 1961 [311], p. 117.

5/ Brandenberg, 1963 [80], p. 57.

6/ Kennedy, 1961 [311], p. 118.

7/ Figures 4 and 5.

and Youden's indexes to ACM papers (1963 [659] and [660]) illustrate single-column formats that alleviate this problem by extending the title line to 103-106 characters, exclusive of the identification code. Youden has calculated that for the titles in the field of computer literature which he analyzed 30 percent of the titles would have been truncated in 60-character title line formats, but that only 2 percent would have been chopped by 103-character title length limits. 1/

A second disadvantageous effect of machine production requirements in most KWIC indexes is the tedious sequential scanning necessary because of the unbroken organization of the page format and the long blocks that occur for frequently occurring word entries. Doyle (1959 [168], 1961 [166]) has investigated this problem of block length and suggests either that alphabetization be carried out to the words following those in the indexing window or that the entries in the block be permuted also in a second-order cycle. The latter suggestion has the advantage of facilitating any two-term coordinate indexing type of search, "because one can now look up directly any pair of subject words, regardless of whether or not they occur adjacently in a sentence." 2/

Redundancy in KWIC indexes, which aggravates the sequential scanning and the long-block fatigue effects, is in large part the result of difficulties in establishing the most appropriate bounds for exclusion or "stop" lists. We have previously distinguished machine-generated indexes of the derivative type from certain of the machine-compiled indexes primarily on the basis that in the first case, the criteria for determining the significance of the keywords to be used as the index access points are applied automatically during the machine processing, even if the selectivity so achieved is only "negative selectivity." 3/ The amount of index entry redundancy, of too many entries and of irrelevant entries is, in simple KWIC indexing, a direct function of the length and contents of the stop list.

In Luhn's original proposals for both KWIC and other types of automatic indexing, he pointed out the importance of the rules which must be established in order to differentiate the significant words from the nonsignificant. He says, for example:

> "Since significance is difficult to predict, it is more practicable to isolate it by rejecting all obviously nonsignificant or 'common' words, with the risk of admitting certain words of questionable value. Such words may subsequently be eliminated or tolerated as 'noise'. A list of non-significant words would include articles, conjunctions, prepositions, auxiliary verbs, certain adjectives, and words such as 'report','analysis', 'theory', and the like." 4/

1/ W. W. Youden, 1963 [458], p. 331.

2/ Doyle, 1961 [166], p. 13.

3/ Artandi, 1963 [20], p. 15.

4/ Luhn, 1959 [381], p. 289.

Interesting variations are to be noted in the current practices of using stop lists. Some lists are quite short, and others extend to several thousand words. Parkins reports that a mere 14 words on the stop lists used for B. A. S. I. C. are responsible for 80 percent of the title lines that need not be printed, but that their original list of 200 stop words grew quite rapidly to more than 1,000 now in use. [1] Chemical Abstracts Service representatives reported in 1962 an initial list of about 1,000 words which dropped to 300 at one time and then was increased again to the original level. [2] Using a stop list of 82 words eliminated 30 percent of a 42,000-word corpus of internal reports at the System Development Corporation, (Olney, 1961 [456]).

Critical questions in the establishment of stop lists relate to the problem of balancing the economics of the number of title lines to be printed and to be subsequently scanned against the loss of retrieval effectiveness if certain words are omitted from the search entry positions. How this balance should be achieved may vary from one subject field to another and between different organizations. In several regularly published KWIC indexes, the actual list used to exclude the presumably nonsignificant words is printed so that the user can check before proceeding to actual search. Williams has suggested that each excluded word be listed once, in its proper alphabetic place in the index, if it occurs in the titles of the particular set of items being indexed. [3]

In general, however, not enough is yet known about the requirements of particular subject fields and particular types of organization to arrive at the most effective compromises in establishing exclusion lists for keyword indexing. Noting that stop lists in actual use vary from only a few function words such as prepositions and conjunctions to lists several hundred words long, Brandenberg points out that:

> "At the present state of the KWIC indexing art the selection of stop words appears to be largely arbitrary and a comparison of half a dozen stop lists shows that they have about two dozen words in common." [4]

Kennedy and Doyle both specifically suggest that more research on the contents and effects of stop lists is necessary, (Kennedy, 1961 [311], 1962 [310]; Doyle, 1963 [162]), but Kennedy points out the ease with which the machine programs themselves can be used for modification of the lists. [5]

---

[1] Parkins, 1963 [466], p. 27.

[2] F. A. Tate, discussions at seminar on the word and vocabulary byproducts of permuted title indexing, Biological Abstracts headquarters, October 8, 1962.

[3] T. M. Williams, discussions at seminar on word and vocabulary byproducts of permuted title indexing, Biological Abstracts headquarters, October 8, 1962.

[4] Brandenberg, 1963 [80], p. 57.

[5] See also Clark, (1960 [123], p. 459), who suggests: "It is very probable... that the cut-off points [for most common, for very infrequent, words] will have to be adjusted to the material we actually use. The effect on the process of such factors as style, size of text, the complexity of the subject matter, and the like, is as yet not clearly seen. The collection of large amounts of text and their analysis will undoubtedly be the best way of determining the effects of these variables."

Some of the reasons for keeping stop lists short, however, may reflect unnecessary programming difficulties. Turner and Kennedy have reported that in the SAPIR system a title word is compared only with the group of nonsignificant words that have the same number of characters, in order to reduce the machine time required for the exclusion list search. [1] Skaggs and Spangler give an account of an exclusion list system developed for general text processing as follows:

> "A representative form developed by General Electric is composed of three groups of words, high frequency, special and standard. The high frequency words (25) occur most frequently in English text. A compression of approximately 35 percent will occur for most kinds of text when these 25 words are deleted. The special words are derived from the particular body of text being processed. The composition of this group is left to the program user. Normally the words for this group are selected by making an Editing list in alphabetical sequence. The words appearing in the index position on the preliminary listing are then reviewed.
>
> "Standard words are words that occur with a relatively high frequency in most types of text and therefore are appropriate for a general purpose screen. In the GE program, 375 words are used in this group.
>
> "To minimize computer processing time, it is desirable that words in the Exclusion Dictionary be arranged in approximate order of their frequency of occurrence." [2]

It should be noted, however, that in most cases stop list searches can be programmed in the form of so-called "logarithmic", "partitioning" or "bifurcation" searches in which the number of machine operations required is only $\log_2 N + 1$, where N is the number of words in the list.

The more words excluded, the fewer the title entry lines that must be included in the final index. This is a factor involving first of all the user in the sequential scanning he must do, where, as Coates has remarked, the retrieval effectiveness is usually in inverse proportion to the amount of such scanning required. [3] Secondly, longer stop lists help to minimize the long block problem, since it is obviously the most frequently occurring title words that have not been excluded that cause the longest blocks of entries.

---

[1] Turner and Kennedy, 1961 [614], p. 7.

[2] Skaggs and Spangler, 1963 [557], p. 29.

[3] Coates, 1962 [134], p. 430.

The important economic factor, however, is the total number of lines to be printed in the index, which is directly reflected in page costs. The effects of page costs, in turn, engender compromises in printing quality, such as page format and size of type. These are among the serious unresolved problems that affect user acceptance of KWIC indexes and involve questions of format, legibility, character sets, and size of the index.

In general, however, in the present state of the art of KWIC indexing, the consensus seems to be that of qualified praise, especially for the early announcement and dissemination applications. The KWIC index is recognized as responding to a definite need,[1] as having merit for fields in which more conventional indexes do not exist as well as for current awareness searching,[2] as receiving excellent response from users "because they can take a handy booklet, sit down at a table and look under the words they know and use, and which they expect other engineers to use in titles." [3] Bernier and Crane, after considering comparative effectiveness data for subject as against word indexing, come to the following conclusions:

> "Title lists keyed by words have value for quick distribution and fast use since time is often a very important element in the obtaining of information. Such lists do not serve adequately for thorough searching. ... A title concordance may be more useful than would seem from the ... data on index entries. However, it must obviously be incomplete, must have many unnecessary entries, and would not prove suggestive enough to users who lack background in the subjects sought." [4]

Additional benefits can quite readily be obtained by taking advantage of the bibliographic information once it is in machine-readable form to provide selective KWIC indexes (Balz and Stanwood, 1963 [28]; Black, 1962 [65]; Carroll and Summit, 1962 [102]) machine retrieval of item citations by specified keywords. (Kennedy 1961 [311]) and selections of items geared to a Selective Dissemination of Information System (Barnes and Resnick, 1963 [36]; Balz and Stanwood, 1963 [28]). Gallianza and Kennedy at the Lawrence Radiation Laboratory, for example, report as being under development programs for the IBM 1401 and 7090 computers which will combine KWIC type indexing features with the logical search operators "AND", "OR", and "IF" in order that users may specify subject searches in ordinary English language terms.[5]

[1] Clapp, 1963 [122], p. 7.

[2] Markus, 1962 [394], p.19.

[3] Black, 1962 [65], p. 316.

[4] Bernier and Crane, 1962 [56], p. 120.

[5] National Science Foundation's CR&D Report No. 11, [430], p. 42.

3.2 Modified Derivative Indexing

Some of the more obvious of the disadvantages of KWIC indexing techniques can be reduced if not eliminated by a variety of human and machine procedures. These include augmentation of titles to provide additional clues to subject aspects, manual post-editing, and synonym reduction through such devices as thesaurus lookups.

The ink was scarcely dry on the first issues of a KWIC index before a number of suggestions for improvements, modifications, and augmentations were proffered in the literature. In fact, both Luhn and Baxendale considered various possible refinements in their original proposals. The first systematic review of work in the field of automatic extracting--whether to produce indexes or abstracts, or both--was made by Edmundson and Wyllys in 1961 [181]. They covered not only the KWIC type indexes as such, but also modifications suggested by Baxendale, Luhn, Oswald and others, and they themselves advanced a number of additional possibilities. Of the various modifications and refinements that have been suggested, the most obvious is that of title augmentation.

3.2.1 Title Augmentation

The machine-prepared index that was probably the first to go into productive operation is actually one involving title and subject indicators rather than pure keyword-from-title permutations. The CIA project, beginning in 1952, is based upon manual pre-editing of the titles themselves, with the words to be picked up as index entries being underlined. In addition, it involves assignment of other words, descriptors or terms from a hierarchical classification schedule to indicate additional access points (Veilleux, 1961 [624].

In later KWIC type indexing, the possibilities of improving effectiveness by pre-editing or post-editing to modify and expand titles have been suggested and explored by a number of investigators. The semi-automatic indexing reported by Janaske adds descriptive words or phrases in parentheses at the end of titles and uses them as additional indexing points (Janaske, 1962 [299]). At Biological Abstracts Service, improvements have been obtained (without sacrifice in the speed desired in order to index 5,000 abstracts twice a month) by title supplementation as well as by an improved stop list and by post-editing word divisions and word recombinations. [1] Titles for each of two 12,000-item bibliographies in the field of radiobiology are reported as being edited considerably before KWIC type processing. [2] Other examples of modified derivative indexing based on title augmentation include Chemical Patents [3], the Applied Physics Letters indexing project at Oak Ridge National Laboratory, which provides for an author-prepared form to describe features of property and method not covered in the title, [4] and the KWIC Index to Neurochemistry ([420]).

---

[1] Parkins, 1963, [466], p.27.

[2] Davis, 1963 [150], p.238.

[3] See Markus, 1962 [394], p. 19, and ref. [662].

[4] Connolly, 1963 [136], p.35.

To some extent, however, the use of human editors to improve the product of KWIC type indexing defeats the initial purpose of a quick and purely clerical or mechanical process. Thus, Dowell and Marshall argue:

> "... The basic permuted-title index can be substantially improved by editing and rewriting the titles before they are submitted to the computer. ... But this of course, destroys the great advantage claimed for the permuted title index, 'that it is a purely clerical process'. Intellectual effort has entered the picture again and we are back where we started." [1/]

In the extreme case, the re-introduction of intellectual effort is in effect the re-introduction of conventional human indexing, with the machine's role limited to that of compilation, as in the case of the "notation-of-content" statements prepared for NASA's STAR System (Slamecka and Zunde, 1963 [561]; Newbaker and Savage, 1963 [430]).

Kennedy suggests instead, therefore, that the augmentation might be accomplished by the authors themselves. However, it may then be pointed out, as by Bernier and Crane, for example, that the supplementation of titles before publication in order to provide suitable additional indexing words would be "awkward, space-consuming and difficult". They continue:

> "It would call for the attention of index experts at the manuscript stage, which would delay publication and expand the total indexing effort. Furthermore, good, thorough indexes are based on the full information of abstracts and papers, not on their titles only." [2/]

An alternative method for title augmentation to improve the quality of KWIC indexing is therefore to establish procedures for machine selection of significant words from more of the text than just the titles alone. In fact, Luhn himself did not limit his technique as originally proposed to titles only but indicated that the process could be performed at various levels: title, abstract, or full text. [3/] In the 1958 permuted index to the ICSI preprints, entries were derived from titles, author's names, author affiliations, headings within the paper, figure and table captions, and sentences and phrases taken directly from text. [4/] Combinations of human and machine procedures based on sentences and phrases selected from text are described by Herner who cites a two-fold advantage: "First, it is not wholly dependent on the informativeness or lack of informativeness of titles and bibliographic citations, and, second, it affords a greater depth of analysis than is generally possible where titles or bibliographic descriptions alone are used." [5/]

[1/] Dowell and Marshall, 1962 [159], p. 324-325.

[2/] Bernier and Crane, 1962 [56], p. 117.

[3/] Luhn 1959 [381], p. 289.

[4/] Citron, et al, 1958 [120], p. i.

[5/] Herner, 1963 [264], pp. 1-2.

Taking more text as the basis for automatic derivative indexing adds, of course, the problems and costs of keystroking additional input material. At the same time, most of the major problems of scatter of references, synonymity, redundancy and exclusive reliance on the author's own language and terminology not only remain but may quite probably be intensified. The problems of establishing suitatle rules for selection of significant words are aggravated, not only by the far larger number of different words to be processed, but because of unresolved problems in effectively relating length of index and depth of indexing to the length of the document.[1]

There are, however, a number of practical suggestions by which machine augmentation of titles might be accomplished. First is the invariant selection of words that are capitalized, other than those that begin a sentence.[2] As Wyllys points out, this type of selection criterion would emphasize proper names, and these in turn might be particularly valuable clues, especially in a military intelligence situation.[3] It has also been suggested that the selection criteria should depend on particular pre-specified contexts, such as being preceded by the words: "the results were...,", "in conclusion...", and the like.

A second type of machine selection procedure is the converse of the exclusion or stop list, namely, an inclusion list or dictionary which may involve especially significant words for a particular subject matter area or words that are of importance to a particular organization. In the discussions of the Area 5 ICSI papers it was remarked:

> "Another complication is that mechanized indexing finds in a paper what was important to the author. What happens if there is something in the paper not important to the author but of importance to the indexer? One possibility is to have a list of words and phrases expressing the interests of a particular collection, which the machine looks for in the papers. If this word or phrase occurs even once, it should be picked up as an indexing term."[4]

---

[1] See, for example, Wyllys, 1963 [653], p. 22.

[2] See Luhn, 1959 [371], p. 52; [384], p. 8.

[3] Wyllys, 1963 [653], p. 15.

[4] See Ref. [578], p. 1263. See also, among others, Luhn, 1959 [371], p. 52: "Just as common words have been eliminated by look-up in a special index, certain essential words may be looked up in another special index for the purpose of listing them under any circumstances".

This approach to the selection problem can be combined with other devices, as in the "Selective Dissemination" system described by Kraft in which keyword extraction indexing is applied to abstract, title, author's name and manually assigned index terms, after processing of all input material against both "in" and "out" dictionary lists. [1]

The use of abstracts rather than full text as source material makes the selection criteria problems somewhat less severe. In addition, there is evidence to suggest that the abstract does contain much of the significant information that would normally be indexed and the text of the abstract is therefore a fertile field for title augmentation. In experiments conducted by Slamecka and Zunde on the comparison of indexing terms manually assigned with the occurrences of the names of these terms in abstracts used in NASA's STAR system, it was found that 80.4 percent of the assigned terms were contained in the abstracts. [2] Swanson, on the other hand, suggests that, at least for short articles having homogeneous subject matter, title and first paragraph "are nearly as good as full text." [3]

A combination inclusion-exclusion list system may involve prior "weighting for relevance" of words that are judged by human analysts to be significant for purposes of search and retrieval, as suggested by Swanson, for example:

> "The computer first separates those words which are important for purposes of information retrieval from those which are unimportant. This is accomplished by means of looking up each word in an alphabetized word list with which the computer is furnished. Each word in this word list carries a 'weight' which reflects an estimate of its importance for retrieval purposes. Words of zero weight are completely unimportant and discarded by the computer for indexing entries." [4]

Continuing work at Thompson Ramo-Wooldridge on automatic indexing methods includes further investigation of assignments of relevance weight estimates to words and phrases, (1959 [490] and [491], 1963 [602]).

3.2.2 Book Indexing By Computer

For internal indexing, that is, the subject indexing of the contents of a single book or report, automatic indexing experiments are usually directed toward the processing of full text, with use of stop lists of various lengths. The work of Artandi for her doctorate at Rutgers in indexing of a book by computer programs (1963 [20] and [22]) is an example of such modified derivative indexing. Specifically, Artandi's method involves:

---

[1] Kraft, 1963 [334], pp.69-70.

[2] Slamecka and Zunde, 1963 [561]. In addition they report (p.139) that a large number of the terms not found were "either broad, general terms (i.e., 'device') or generic level concepts of terms contained in the abstracts."

[3] Swanson, 1963 [580], p. 1.

[4] Ibid, p. 1.

(1) Establishment of a list of key terms appropriate to a given subject area to be used as an inclusion list for word extractions from text.

(2) Application of an appropriate syndetic apparatus to be used in the compilation and ordering of the index entries.

(3) Means for the automatic selection of index entries other than those on the pre-specified inclusion list, especially for the selection of proper names.

The text used by Artandi for her study consisted of a 59-page chapter on halogens from J. W. Mellor's Modern Inorganic Chemistry. This text was keypunched with special tags being assigned to indicate the page numbers and the incidence of capitalized words in the text. Text words greater than three characters in length were first checked against the inclusion dictionary of "detection terms". There was, in addition, an "expression term" dictionary which constituted the vocabulary of the final index and in which a given expression term might or might not be identical with the corresponding detection term. Cross-references were supplied by a program routine which checks the index term list against a list of expression terms with their detection terms grouped under them and which compiles cross-reference entries, one for each detection term associated with an expression term appearing on the index list.

For her experimental corpus, Artandi's program developed 363 page references, 138 different index entries and 35 cross-references. She compared these results with those obtainable by conventional human indexing with respect to the factors of heading density (ratio of number of entries to number of words in the book), entry density (ratio of the number of page references to the number of pages), and distribution (ratios of entries for chemical compounds, proper names, and subject entries to the total number of entries. No indexing errors were found in the computer-generated index for a 5 percent random sample of the pages of the corpus, but five omissions were found in the machine indexing of these sample pages. Artandi concluded, however, that although the quality of indexing appeared favorable, the costs, which approximated $1.50 per page indexed, were impractically high.

Book indexing by computer has also been investigated by Maloney, Dukes, and Green at the Army Biological Laboratories, Fort Detrick, Maryland.[1/] Input is based on the by-product paper tape generated when the manuscript is typed on a tape typewriter. The paper tape is in turn converted to punched cards which are then processed by a UNIVAC SS-90 II computer in an editing run that deletes unrecognizable codes and then stores page,

---

[1/] C. J. Maloney, private communication. A report by C. J. Maloney, J. Dukes, and S. Green, "Indexing reports by computer" is in process of preparation for publication.

line, sentence number and other reference identifications. After re-processing against a stop list of common words, all other words in the edited text are selected as candidate index entries, these are then sorted into alphabetical order with subsequent printout giving each word occurrence followed by the entire sentence which contained it and the page and other location identifications. This computer output is then post-edited manually not only to eliminate trivial entries but also to normalize terms and phrases used.

3.2.3 Modified Derivative Indexing - Baxendale's Experiments

As has been previously noted in the introduction to this report, the name of Phyllis Baxendale together with that of H. P. Luhn is generally accorded credit for pioneering efforts in the entire area of automatic indexing. Baxendale in particular is generally credited with the first actual experiments in modified derivative indexing. In investigation beginning in the late 1950's, she has explored not only statistical approaches to automatic selection of index terms (based for example on word frequencies) but also the use of word pairs, word groups, contextual associations, and in particular the subject-indicating clues of prepositional phrases (Baxendale, 1958 [41], 1961 [40], 1962 [42]; Becker, 1960 [44]; Edmundson and Wyllys, 1961 [181]).

Baxendale began by considering the patterns of scanning that humans typically use to select "topic" sentences, phrases and words, and she then proceeded to simulate by computer program the selection of phrases consisting primarily of nouns and modifiers. In her first experiments, (1958 [41]) she used two methods of automatic selection. In the first procedure, words serving the grammatical functions of pronoun, article, auxiliary verb, conjunction and the like, were deleted by stop list lookup. Frequency count statistics were then derived for the remaining words. In her second procedure, the computer was programmed to select prepositional phrases from text and to use the four words succeeding the preposition as index entries unless an additional preposition or a punctuation mark is first encountered.

In later experiments, Baxendale has explored possible grammatical models "which would select all and only nouns or adjective-noun combinations". [1] Taking as an initial corpus a sample of document titles, rules were devised to reject for human analysis titles with question-marks and the like, to eliminate numeric information and single symbols, and to segment the title into its component clauses and phrases by the detection of commas, periods, and similar clues. By list lookup, certain words are identified as capable of serving the syntactic functions of being quantifiers, prepositions, or clause introducers. Special subscripts are then assigned to these words and the subscripts are examined by machine to provide further segmentation; to delete quantifiers, auxiliary verbs, or words ending in "ed" or "ing" and preceded by an auxiliary verb, and to determine relationship functions between the remaining, presumably substantive, words.

Still other work by Baxendale has been directed toward the development of frequency of co-occurrence or textual association of candidate indexing terms. She reports as follows:

---

[1] Baxendale, 1961 [40], p. 209.

> "[In the frequency matrix]. . the diagonal elements . . . give the total frequency of an index term and the off-diagonal gives the frequency of co-occurrence of two terms. The diagonal of the 'context' matrix represents that portion of the total vocabulary with which an individual term has been coordinated, and the off-diagonal the extent to which two terms have common context. . . Such matrices give a basis for examining the extent to which terms are generic or specific within the context of the collection of documents. One can speculate that terms occurring with high frequency and wide context, i. e., with frequencies distributed amongst all or nearly all off-diagonal elements of the matrix are of such broad connotation as to be indifferent discriminators of content . . . The frequency and context matrices can again be used to determine the modifiers with which they can most meaningfully be coupled for the collection of documents being considered." 1/

Finally, Baxendale notes that on the basis of her studies it should be possible to select quasi-subject headings based on frequency counting criteria, but then to order the remaining vocabulary of selected terms according to contextual measures of association which are semantic, syntactic, or statistical in nature. Experimental results for a collection of 1, 500 documents included semantic associations between "searching" and "retrieval", syntactic associations of "machine" or "literature" with "retrieval", and the apparently misleading association of "metal" with "retrieval" which, however, had statistical significance within the particular document sample. 2/

Other investigators who have explored noun-adjective clues for selection include Anger, Chonez, Langleben and Shumilina, and Swanson. Anger looked for relationships indicated by syntactic dependencies or by noun-adjective and adjective-adverb linkages, and gave in an appendix a suggested program for phrase inversions. 3/ Chonez has described a computer program which by recognizing "separating" words, especially prepositions, and applying "pseudo-grammatical" rules compiles an index to English language items in the fields of ionized gas physics and thermonuclear fusion. It is claimed that:

> "The subject index thus prepared is similar in presentation to Luhn's KWIC indexes, but is fundamentally different in conception and is in fact intermediate between. . . (this) . . . and the conventional alphabetic subject indexes." 4/

Langleben and Shumilina are concerned with machine-aided procedures for translation from natural language materials to an intermediary or documentation language.

1/ Ibid, pp.215-216.

2/ Ibid, pp. 216-217.

3/ Anger, 1961 [15], pp. III-6 ff.

4/ Chonez, et al, 1963 [119], p. 31.

They indicate, for example, that the preposition "from" serves as a key for the treatment of two nouns connected by it. [1] Swanson, describing research project progress at Ramo Wooldridge as of 1960, reported to the National Symposium on Machine Translation with respect to multiple meaning problems as follows:

> "We are also investigating the possibility of discovering semantic attributes of words based upon certain automatically recognizable statistical features of the context. Our initial endeavor in this direction has been to attempt to discover a classification system for nouns based upon their frequency spectrum of categories of modifying adjectives, these categories being automatically recognizable." [2]

3.3 Derivative Indexing From Automatic Abstracting Techniques

While Baxendale's work has had certain points in common with automatic abstracting or extracting processes, particularly in the use of word frequency statistics and the consideration of possibilities for first selecting topic sentences, her major interests in this area have been in automatic indexing as such, rather than in machine selection of sentences from text to serve as an automatic extract or derivative abstract of the document. Much of the machine processing to date of full text for documentation purposes, however, has had the latter goal as the principal research objective.

As we have previously noted, the subject of automatic abstracting or auto-condensation is not in itself a primary concern of this survey. Nevertheless, the significant words occurring in the abstract of a document, whether generated by man or by machine, are obviously good candidates for indexing terms. Moreover, it has been strongly suggested that the questions of using positional, editorial, and syntactical clues in order to improve automatic indexing techniques will profit by research that is being done in both automatic extracting procedures and in other types of linguistic data processing based upon full text. [3]

3.3.1 Auto-Condensation and Auto-Encoding Techniques of H. P. Luhn

Although Luhn's work in the field of documentation aided by machine has had its best known and most popular acceptance with respect to the KWIC index proper, even more provocative possibilities lie in the development of some of the auto-condensation and auto-encoding techniques which he also proposed, especially for full text processing. In this area, although he himself has also suggested a variety of possible improvements and refinements, the actual experimental work done by him and by his associates has mostly been done on the basis of word frequency statistics.

---

[1] Langleben and Shumilina, 1962 [347], p. 109.

[2] Swanson, 1961 [585], pp. 391-392.

[3] See, for example, Wyllys, 1963 [653], p. 7.

Considering first the most frequently occurring words in a given text as too common to be subject-indicative (those usually stopped or purged by a suitable exclusion dictionary or stop list, for example) and next the least frequent words as being rarely topical in a content-revealing sense, Luhn settles upon a middle range of frequency of word occurrence as the basis for his auto-condensation processes. The actual frequency counts are computed, together with indications of page, line, and occurrence within the same sentence. When this has been done for the complete text, each individual sentence is then checked for the "score" of relatively high frequency words occurring in it, and sentences with the highest scores are then automatically selected, in textually-occurring order, and are printed out as an abstract, more properly an extract, of the document.

The automatic encoding of documents may be achieved either by taking the high ranking words of the selected sentences or by selecting the highest ranking of the words in the entire document as index entries. Luhn typically justifies these procedures as follows:

> "Of various automatic procedures for deriving typical patterns for characterizing documents, the systems here proposed are based on operations involving statistical properties of words ... It is held that the more often a certain word appears in a document the more it becomes representative of the subject matter treated by the author. In grading words in accordance with the frequency of usage within a document, a pattern is derived which is typical of that document and unique amongst all similarly derived patterns of a collection of documents. It is proposed that the more similar two such patterns are the more similar is the intellectual contents of the documents they represent...
>
> "... The creation of an encoding pattern may consist of listing an appropriate portion of the words ranking highest on the word frequency list derived from a document. Experiments conducted so far on documents ranging in size from 500 to 5000 words have indicated that word patterns consisting of from ten to twenty-four of the highest ranking words furnish adequate discrimination and resolution for retrieval, sixteen such words being a likely average." [1/]

At Wright-Patterson Air Force Base an automated information selection and retrieval system has been developed jointly by Air Force and IBM personnel (Gallagher and Toomey, 1963 [205]). It involves both auto-indexing and auto-abstracting techniques following the Luhn word-frequency-counting techniques. Pre-editing is applied to demarcate fields (e.g., title, author) and to flag certain text words, particularly proper names, for special treatment. Special treatment, over and above the frequency-based selection score, is also given to words in the title field.

On the abstracting side, modifications to the original Luhn formula involve segmenting sentences in terms of strings of both high and low valued words separated by either periods or continuous strings of low valued words, on the assumption that long consecutive strings of low value words should weight negatively. The automatic extract consists of the highest ranking 20 percent of the sentences subject to the restriction that no less than 7 and no more than 20 sentences should be selected. On the indexing side, the investigators report:

[1/] Luhn, 1959 [371], p. 47.

> "As it is currently run, the auto-indexing program selects about one word in ten as a keyword in articles of three thousand words or less. In articles longer than three thousand words it tends to pick about one word in fifteen. This high incidence of keywords naturally increases the amount of noise results returned by the query program, although good search strategy cuts them down considerably." [1]

As of October 1963, the system was reported to be fully operative although not as yet extensively tested in actual use. Gallagher and Toomey give illustrative auto-extract results on two tested papers, one being Luhn's own "Automatic Creation of Literature Abstracts". They give comparative results for manual versus machine selection of keywords as index or search terms with 88.6 percent agreement, the human indexers having selected, in 6 tests reported, 132 words and the machine method 117. Modifications under consideration include pre-edit flagging of terms in author and cited-reference fields for special weighting, setting the length of the abstract as a function of the total number of words in an item, and, in the search program, generating additional search terms by means of association factor techniques such as those suggested by Stiles.

To the basic approach of straight-forward word frequency counting, Luhn himself has suggested that improvements might be obtained from considering closely adjacent words, [2] word pairs, [3] and reference to vocabularies specific to a given field. [4] Other possibilities are capitalized words and lookup against an inclusion list. He also suggests:

> "If certain words could be given in their relationships to other words, more specific meanings may be identified by such combinations. These relationships may range from the mere co-occurrence of certain words within a phrase or sentence to the combinations of specific parts of speech." [5]

Various investigators have proceeded to explore these and other possible improvements, including incorporation of relative frequency information, use of information about distances between high-ranked significant words, word pairs and word n-tuples, and other devices to improve detection of significant clues to subject content. Representative examples of such work will be discussed below. In addition, investigators abroad have developed modifications to the basic Luhn word frequency approach which appear to be necessary when it is applied to languages other than English. [1]

---

[1] Gallagher and Toomey, 1963 [205], p. 51.

[2] Luhn, 1959 [384], p. 10.

[3] Luhn, 1962 [373], p. 11.

[4] Luhn, 1959 [384], pp. 8 and 10.

[5] Ibid, p. 5.

Thus, for example, Purto reports various investigations conducted by V. A. Argayev and V. V. Borodin and by himself with respect to Russian language documents. [2] Purto notes first that the Luhn method as applied to Russian language materials selects sentences which, while having the largest "significance coefficients", were not those most essential to the meaning and further that: "an abstract in Russian made by Luhn's method results in a choice of sentences not conveying basic information and not logically connected with each other." [3] The reasons for such failure he attributes to the fact that words with different frequencies are considered equally important within a sentence for sentence selection purposes and to the lack of consideration for semantic and grammatical connectivity between significant words and between sentences. He then discusses several methods for determining connectivity, such as the rule that the sentences most closely connected with each other will be those in which the greatest number of the same significant words occur. [4]

A somewhat different example of difficulties occurring when the basic Luhn technique is applied to material in languages other than English is given by Levery. He describes a study of thirty French texts concerned with the development and manufacture of glass. He reports as follows:

> "While we followed the classical idea that a relationship between the frequency of a word and its significance exists, the fact that we worked with French texts forced us to discount the value of frequency alone.
>
> "French authors generally do not like to repeat the same words, and they vary their vocabulary... It was necessary to combine the frequencies of words with the same meanings or related to the same idea."
>
> "A dictionary of synonyms was constructed... (and) different versions of the same word had to be regrouped." [5]

---

[1] Note, however, that in the automatic abstracting program at Thompson Ramo-Wooldridge, small-scale experiments suggest that automatic abstracting is as feasible for other Indo-European languages as for English, (1963 [603], p. ii). Also, at the Centre d'Etudes Nucléaire Saclay, automatic extraction experiments are being applied to texts both in French and other languages, see National Science Foundation's CR&D report No. 6, [430], p. 20.

[2] Purto, 1962 [484]. He refers to a report "The problem of automatic abstracting and a means of solving it", by Argayev and Borodin, apparently available only as a typescript dated 1959.

[3] Ibid, p. 3.

[4] Ibid, pp. 3-4.

[5] Levery, 1963 [359], p. 235.

3. 3. 2 Frequencies of Word n-tuples - Oswald and Others

The first alternative to the basic Luhn word frequency approach in automatic abstracting techniques to be actively explored was apparently that of Oswald and his associates. (Oswald et al, 1959 [459]; Edmundson et al, 1959 [180]). Like Baxendale, Oswald was interested in word pairs and word groups, particularly compound-noun and adjective-noun compositions, as more revelatory of meaning than single words. Unlike Baxendale, however, he was interested in the word group itself as selection criterion, whereas she had used word group or phrase clues for the selection of (usually) single indexing terms. Differences between their two approaches, both representing very early efforts in the field, are summarized by Edmondson and Wyllys as follows:

> "Oswald's experiment in automatic abstracting differs from Luhn's and Baxendale's techniques in that it combines the notion of significance as a function of word frequency and the notion of significance as a function of word groupings, by employing juxtapositions of significant words as the basic unit for measuring the importance of a sentence...
>
> "It may further be observed that Baxendale's exhibited indexes are made up of single words rather than word groups, in spite of the strong case she makes for using groups...
>
> "Baxendale's work is concerned solely with the automatic construction of indexes; she does not extend her treatment of word significance into the area of automatic abstracting." [1]

Oswald's "multiterms", however, were intended to overcome, in the areas of both automatic indexing and automatic abstracting, at least some of the difficulty that concepts are often expressed in compound nouns, word pairs, and longer groups of words consisting of n-tuples of substantive words or of phrases. The result of considering both word frequency and word-group frequency is that in Oswald's selection-groups it is usually the case that only one word of the group has an individually high frequency but the co-occurrence feature heightens the significance of the relatively lower frequency words with which it appears. Thus, for automatic indexing, Oswald proposed significant word groups as indexing terms, and his criteria for selection of sentences to be included in machine-generated extracts are similarly based on the number of significant groups in the sentences chosen.

Other investigators who have stressed the importance of word pairs and longer groups as necessary to reflect concepts include Bar-Hillel (1959 [33]), Black(1963 [64]), Clark (1960 [123]), Doyle (1959 [165]), and Salton (1963 [519]). Doyle says succinctly that "when a phrase, or some other aggregation of words, stands for a single idea, its frequency in a document ought to interest us more than the frequencies of its component words." [2] Salton considers it desirable to use word groups rather than individual words for purposes of identifying document contents and to use data on the joint occurrence of words in the same sentence or similar contexts as grouping criteria. Clark points out in particular that the use of ordered pairs and longer sequences of words to express a single concept may be highly characteristic of the special technical language used in a specific subject field, and notably those of the social sciences. 1/

---

[1] Edmundson and Wyllys, 1961 [181], pp.231-232.

[2] Doyle, 1959 [165], p. 11.

Others who have explored word n-tuples as selection criteria for automatic extraction operations include such investigators as Szemere, Levery, and Yakushin. Szemere reports an investigation of 39 Swedish patent specifications in the field of switching circuits looking for significant word-pairs, with emphasis on noun-adjective combinations (1962 [591]). The objectives of a project headed by Levery at IBM - France have been reported as follows:

> "A series of experiments is planned in the fields of automatic indexing of technical texts and technical vocabulary analysis.
>
> "A statistical method will be tested to determine the degree of closeness in meaning of words. The method will consist of studying the pairs of words which appear together in the majority of texts and calculating a coefficient of correlation from the frequencies. Such work will result in a standard list of notions frequencies for a particular kind of information.
>
> "Starting from this list, new experiments will be made so as to obtain a list of keywords representing each text. The method will use statistical comparison between the distribution of frequencies of notions contained in a text and the standard distributions obtained for the entire corpus." 2/

Yakushin(1963 [654]) develops a variation of the word-pair principle in which he looks for those pairs where the words are, or suggest, names of objects, such as "table-leg". He suggests, further, that so-called "basis nouns" can be established for a given scientific field and entered into an inclusion dictionary, which also contains codes for the lexical classes to which the word can belong and codes for determining whether or not the word can join with another as a "basis term". Machine routines are then suggested to develop whether or not given terms are jointly part of the same text, whether one textually precedes another in a given text, whether or not there is a "nomenclator" pair. Depending upon the frequency of occurrence of identical or semantically related nomenclator constructions, it is claimed that subject concepts can be detected. That is:

> "The method is founded on the finding in a text of so-called basis terms, established by list, and of the words which explain them. These explanatory words, which in different contexts refer to one basis term, are grouped and ordered according to definite rules into a subject concept." 3/

1/ Clark, 1960 [123], p. 460.

2/ National Science Foundation's CR&D report no. 11, [430], p. 118.

3/ Yakushin, 1963 [654], p. 16.

3.3.3 Relative Frequency Techniques - Edmundson and Wyllys, and Others

The first comprehensive critique of word frequency approaches to automatic extracting and indexing was undoubtedly that of Bar-Hillel (1959 [33], 1960 [34]), followed closely by Edmundson and Wyllys (1961 [181]), who themselves have experimented with various alternative or improved methods for obtaining measures of word significance by statistical analysis. These critics have been in agreement both on many points of specific criticism and on suggested possibilities for amelioration of observed difficulties, especially in terms of considering relative word frequencies within a particular subject field. In addition, several other investigators independently proposed a relative frequency approach at about the same time. [1]

Some typical expressions of opinion on the importance of relative frequency criteria are as follows:

> "Let me propose here a system of auto-indexing which, to my knowledge, has never been publicly proposed before in this form and which seems to me superior to any other system I have heard of ... Assume that ... we are given a list of the average relative frequencies of all English 'words' ... It would then be possible, for any given document, to rank-order all the 'words' occurring in this document according to the excess of their relative frequency within the document over their average relative frequency. By some mechanically implementable standard or other, an initial segment of this list is selected as the index-set." [2]

> "Very general considerations from information theory suggest that a word's information should vary inversely with its frequency rather than directly, its lower probability evidencing greater selectivity or deliberation in its use. It is the rare, special, or technical word that will indicate most strongly the subject of an author's discussion. Here, however, it is clear that by 'rare' we must mean rare in general usage, not rare within the document itself. In fact it would seem natural to regard the contrast between the word's relative frequency f within the document and its relative frequency r in general use ... as a more revealing indication of the word's value in indicating the subject, matter of a document." [3]

---

[1] Compare, for example, Kochen, 1963 [327], p. 7: "The idea of contrasting words which occur frequently in a document against the frequency of this word in the background language for purposes of selecting index terms seem to have been suggested first by Bohnert and the author, then described in more detail by Edmundson and Wyllys, and tested empirically by Damerau. Something similar was suggested even earlier by Bar-Hillel." See Bar-Hillel, 1962 [35], p. 418, footnote, with respect to himself, Edmundson, and Bohnert. See also, however, Doyle 1962 [163], p. 388: "Edmundson and Wyllys were probably the first to publicly advocate contrasting word frequencies within a document to word frequencies within a given field and using these relative frequencies as criteria for scoring and selecting sentences."

[2] Bar-Hillel, 1959 [33], pp 4-8-9.

[3] Edmundson and Wyllys, 1961 [181], p. 227.

"We naturally find that the words of greatest interest are those for which there exists the greatest contrast between general usage frequency and local (within the article) usage frequency." 1/

"Luhn has bypassed syntactical analysis by taking advantage of the information content of the most frequently used topical words in articles ... Edmundson et al take a further step in a desirable direction by bringing in information from outside the article being analyzed: words and terms are given greater topical value as the contrast increases between the frequency of use within the article and the rarity of general usage." 2/

"A further refinement of the process of automatic analysis would be the development of special sets of reference frequencies for special fields of interest. This would have two benefits: it would become possible to classify documents as to field, and it would become possible to note the significance of words which are frequent in the document and frequent in a very large reference class c_o of literature (i.e., these words would not be significant with respect to c_o) but which are rare in the special field. For example, the word 'emotion' might be too common in general usage to seem significant, but frequent occurrence of the word would stand out in a paper on electronic circuitry (e.g., of a robot) when compared with its frequency in general electrical engineering literature." 3/

"One of the ... goals is to investigate a relative-frequency approach to the categorization of documents... For this investigation it will be necessary to develop sets of reference frequencies for words used in different subject fields. It was suggested by Edmundson and Wyllys that these sets of reference frequencies, when developed, could be used to categorize a document as belonging to a particular subject-field, by means of measuring the degree of matching (e.g., with the chi-squared test) between the proportional frequencies of words in the documents and the sets of reference frequencies." 4/

Two points in the comments quoted above appear especially worthy of note. The first is that of introducing at least some measure of reference to material other than the individual author's own choice of linguistic expression and specific terms. We shall discuss this factor in more detail in a later section of this report. The second point, derived in part from the first, is the specific suggestion of movement away from purely derivative indexing by machine in the direction of automatic assignment indexing and automatic categorization or classification.

1/ Doyle, 1959 [165], p. 9.

2/ Doyle, 1961 [169], p. 3.

3/ Edmundson and Wyllys, 1961 [181], p. 228.

4/ Wyllys, 1963 [653], p. 10.

Actual experiments in application of relative frequency techniques to automatic extracting processes have been pursued since 1959 by various investigators. Edmundson and Wyllys and Damerau (1963 [148]) were certainly among the first. Edmundson and Bohnert were engaged in experimental investigations at Planning Research Corporation in 1959, [1] and the following year Edmundson, Oswald, and Wyllys worked on the auto-indexing and auto-extracting of the 40,000 words of text contained in nine articles in the subject field of missilery. [2] Wyllys has continued work on relative frequencies (1963 [653]). At the System Development Corporation Doyle, in some of his work, has also explored the relative frequency approach (1961 [161]). An example in Europe is work reported by Meyer-Uhlenried and Lustig, where significant keywords from abstracts are used not only as indexing terms directly, but by means of keyword lists and micro-thesauri can also be used to assign documents to specific subject fields (1963 [417]).

3. 3. 4 Significant Word Distances

Another technique that has been investigated for the improvement of automatic extraction operations based on the statistics of word frequencies is that of distances between significant words. The desirability of attaching greater weight to n-tuples of immediately adjacent words and to the co-occurrences of words within the same sentence has been mentioned previously. Savage, in relatively early work developing some of the initial proposals of Luhn, considered intra-sentence distances between significant words as follows:

> "... The criterion is the relationship of the high-frequency words to each other, rather than their distribution over the whole sentence. Consequently, it seems reasonable to consider only those portions of sentences which are bracketed by high-frequency words and to set a limit for the distance at which any two such words shall be considered as being significantly related ... An analysis of many sentences and many documents indicates that a useful limit is four or five non-significant words between any two high-frequency words." [3]

Doyle has also noted the tendency of words that are in fact highly related in a content-revealing sense to co-occur in the same sentence or as quite direct neighbors. The same investigator has also suggested that word distances can be used to provide "clustering" effects that might, for example, sort out the possibly different topics covered in introductory or background discussions, the main text, and various appendices. [4]

---

[1] National Science Foundation's CR&D Report No. 5, [430], p33; Bar-Hillel 1962 [35], p. 418.

[2] National Science Foundation's CR&D Report No. 6 [430], pp 43-44.

[3] Savage 1958 [521], p. 4. Later related work has included a method for generating auto-extracts which adds to the high-frequency word sentence scores a correction factor for the number of words in gaps between such words. (See Rath et al, 1961 [493])

[4] Doyle 1961 [166], p. 7.

Related research efforts in more general areas of linguistic data processing suggest inter-sentence distances as criteria for the selection of words and word groups in automatic indexing and abstracting processes. In natural language text searching, for example, the work of both Swanson (1960 [587], 1961 [586], 1963 [583]), and of Maron and Ray [1] suggests that limitation of searching to a four-sentence span would eliminate a number of irrelevant responses to search requests specifying the joint occurrence of two or more words.

Swanson's findings indicated that if two words or phrases contained in the search request were found in textual proximity within these limits, they were highly likely to bear a semantic relationship that is what was intended by the requester. Applying the four-sentence proximity criterion, it was found that the amount of irrelevant material retrieved by the text searching system could be reduced by 60 percent without serious loss of relevant information. [2] Black cites the four-sentence proximity criterion and notes further that it might be used also to retrieve only a paragraph or similar small portion of the full text, reducing the amount of material to be read by the user, perhaps by as much as 90 percent. [3]

Artandi, in her book-indexing studies, suggested as a topic for further investigation the possibility that proximity of index term candidates as derived from the same section of the text could serve to improve the quality of the indexing. Since her computer program checks for duplicate potential entries occurring on the same page, this feature could be used for further analysis, on the assumption that the number of occurrences of the same entry for the same page is an indication of the importance of the discussion of the subject on that page. [4]

3. 3. 5 Uses of Special Clues for Selection

Intra- and inter-sentence distances between words are relatively crude examples of clues to selection of words and word-pairs which, because of their implied relationships, may be especially significant for indexing, sentence extraction, or document categorization. They can be quite readily detected by machine, but the implication that physical proximity is a good measure of significant co-occurrence is often false. Other clues which can be detected equally well, mechanically, are those which have to do with position and format.

---

[1] Ray, 1961 [494], p. 92.

[2] Swanson, 1963 [583], p. 9, 1961 [586], pp.298-299.

[3] See Black, 1963 [64], p. 20 and footnote: "The figure 90 percent is derived from experience in previous experiments, wherein the amount of relevant material was scanned and a subjective judgment was formed that the relevant material was actually about 10 percent of the total verbiage retrieved. That is, about 10 percent of each document contained the relevant material; 90 percent of the document was of no relevance but the document as a whole was relevant."

[4] Artandi, 1963 [20], p. 47.

Such obvious positional clues as occurrences of words in titles, chapter or section headings, figure captions, have already been mentioned. To these can be added first and last sentences of paragraphs, [1] or of first and last paragraphs as such. [2] Wyllys observes that other criteria which are detectable in the text by straightforward machine procedures can be based on such features as italicization, capitalization, or punctuation. He notes, however, that such "editorial" criteria vary from journal to journal so that their usefulness would need to be related to the particular practices of individual journals. [3]

Somewhat more difficult for machine implementation, but certainly feasible in the present state of the programming art, is the use of specific semantic or syntactic clues. Here again, Luhn, Baxendale, and Edmundson and Wyllys all anticipate their critics and later investigators. Luhn recognized the fact that in at least some applications the characterization of documents by isolated words alone would fail to provide an effective degree of discrimination. He, therefore, suggested operations to establish word relationships, whether based on co-occurrences or combinations of specific parts of speech. [4] Baxendale clearly uses both syntactic and semantic clues, detectable by built-in table lookups.

Representative suggestions by Edmundson or Wyllys or both as co-authors include the following:

> " . . . We have in mind a glossary or dictionary of perhaps one to two thousand words that act either as cue words which signal the importance of a sentence or as stigma words that signal the insignificance of a sentence for purposes of abstracting." [5]

---

[1] See, for example, Wyllys, 1963 [653], p. 27: "One of the first published studies in automatic document-content analysis, that of Miss Phyllis Baxendale, brought out the importance of the first and last sentences in a paragraph as bearers of a good deal of the content of the paragraph." See also Marthaler, 1863 [399], p. 25.

[2] Compare Swanson, 1963 [580], p. 1: ". . . Some evidence exists to show that for short homogeneous articles title and first paragraph are nearly as good as full text."

[3] Wyllys, 1963 [653], p. 28.

[4] Luhn, 1959 [384], p. 5.

[5] Edmundson, 1962 [178], p. 11.

"The criteria for attributing significance to words ... may be positional (in virtue of their occurrence in titles or section headings), or semantic (in virtue of their relation to words like 'summary'), or perhaps even pragmatic (in the case of names of specialists mentioned in text footnotes, or bibliography ...

"A cataloguer or abstract-writer would naturally give more weight to a technical word that appears in a title, in a first paragraph, or in a summary. A machine can be programmed to do the same. It can be instructed to recognize the title by position and capitalization ... It can place first-paragraph indications... It can test every heading or subtitle for the words 'summary' or 'conclusions' and place a summary indication after each word in the summary paragraphs." [1]

"The statistical criteria ... by no means exhaust the potential clues to the representativeness of sentences. Among other plausible clues are certain words and phrases ... authors use words such as 'conclusion', 'demonstrate', 'disclose', 'prove', 'show', and 'summary' (and related forms of these) with high frequency in sentences that contain concise statements about the topic or topics of the article... The occurrence in a sentence of such a phrase as 'it was found that...', 'the experiment proves...', or 'the central problem is ...' would indicate probably even more sharply than any single word could that the sentence was likely to be highly representative of the topics..." [2]

3.3.6 Recent Examples of Mixed Systems Experimentation

It is quite obvious from the above samples of suggestions for the use of various special clues for automatic extraction, that improved systems will largely depend upon a mixture of means for determining subject-representativeness of words, phrases, and sentences. Many of the clues suggested by Edmundson and Wyllys are continuing to be explored, as mixed systems, at RAND [3] and the System Development Corporation, (1962 [590]), for example. Two specific recent examples of mixed systems experimentation are the automatic abstracting experiment programs at Thompson Ramo-Wooldridge and the work involving detection of first incidences of nouns at the Harvard Computation Laboratory.

The TRW programs to investigate possibilities of computer generation of document auto-abstracts, involving both English and Russian language texts are based upon a combination of four different methods to measure significance and determine representativeness. These four methods are briefly described as follows:

"... The Key method has its source of machine recognizable clues the specific characteristics of the body of the document and is based on a Key Glossary of content words taken from the body of the document.

---

[1] Edmundson and Wyllys, 1961 [181], pp. 227 and 229.

[2] Wyllys, 1963 [653], p. 25.

[3] See National Science Foundation's CR&D report No. 11, [430], pp. 314-315.

> "... The Cue method has as its source of machine recognizable clues, the general characteristics of the corpus that are provided by the bodies of the documents and is based on a Cue Dictionary of function words apt to appear in the body of a document.
>
> "... The Title method has as its source of machine recognizable clues, the specific characteristics of the skeleton of the document, i.e., title, headings, and format, and is based on a Title Glossary compromising those content words found in the title, subtitles, and headings, but excluding certain words of the Cue Dictionary.
>
> "... The Location method has as its source of machine recognizable clues, the general characteristics of the corpus that are provided by the skeletons of the documents and uses a Heading Dictionary of certain function words that appear in the skeletons of documents." 1/

The Harvard work involving detection of the first incidences of nouns as sentence selection and indexing clues is part of a larger-scale program for mechanized information selection and retrieval under the general direction of Salton (1961 [512], 1962 [513], 1963 [514] and [515]). The specific mixed system involving frequency data, syntactic identification clues, and positional criteria is primarily the result of investigations by Lesk and Storm (1961 [577], 1962 [358]). Related work takes advantage of computer techniques for predictive syntactic analysis and automatic dictionary lookup also under development at the Harvard Computation Laboratory (Kuno and Oettinger, 1963 [339], [340], [341]).

The Lesk-Storm experiments have involved investigations where the hypothesis assumed is that the points in a text where the author has first introduced a specific noun or nominal phrase, or where he has used, with higher frequencies, a combination of first-referred-to-nouns, are most likely to be especially indicative sections of text with respect to subject-content representativeness. The assumption is further, that areas in which specific "new" ideas, not mentioned previously in the text, are first introduced is particularly rich in topical-content concentration. 2/

The mixed-system emphasis followed by Lesk and Storm, however, is revealed in the following comments:

> "It is not, of course, apparent that a count of initial occurrences of nouns ... is by itself sufficient to reveal areas of significant information content for purposes of abstracting or indexing. Accordingly, the method suggested here must be used together with other available means, and is not expected to provide by itself an acceptable abstracting algorithm." 3/

In their actual investigations, Lesk and Storm first made manual counts of initial noun occurrences in various sample texts, noting paragraph, sentence, and first incidence-of-word identifications. The computer was then used to carry out three distinctive tasks: (1) calculation of the number of new nouns for each sentence in the text;

1/ Thompson Ramo Wooldridge, 1963 [603], p. 1.

2/ Lesk and Storm, 1962 [358], p. I-6.

3/ Storm, 1961 [577], pp. I-1 and I-2.

(2) computation of functions proportional to the number of initially occurring nouns for each sentence, and (3) the preparation of a normalized graph for initial noun occurrences by plotting the functional values against each sentence in the text.[1] Sentence selection can then proceed by processes to detect "peaks" on the graph, using a relative criterion or weighting function to minimize the effect of high first-noun counts in the beginning sentences of a paper.

Trials were made with a number of different weighting formulas, and the best of these involved the obtaining of moving averages of first-noun counts over several adjacent sentences. A particular formula covering a span of seven sentences gave results that appear to emphasize contextual effects and to reduce the effects of a particular single sentence with a large number of new nouns, such as a listing of proper names. The resulting abstracts are quite lengthy (e.g., comprising 20 percent or more of the original text), and contain some relatively uninformative sentences. The investigators think that the results with respect to satisfactory abstracting are inconclusive but provocative. They also conclude that the possibilities for indexing are more immediately promising: "Most key definitions are retained in the successful summaries, and the vocabulary reflects the topics covered in the texts." [2]

Other examples of mixed-system experimentation, especially involving the use of syntactic and semantic considerations, include the work at the General Electric Computer Department under Spangler, and work by Jacobson and Plath. In the Phoenix laboratories of General Electric, a KWIC type indexing program can be applied both to titles and to running text and a contemplated extension is intended to "generate indexes by means of word analysis, taking into consideration syntactic and semantic aspects of text lines".[3] Jacobson describes rules for machine determinations of same-meaning occurrences of words which may be homographic and for selection of descriptors for indexing simple paragraphs by choosing words occurring at least twice with a high probability of having the same meaning. [4] Plath reports:

> "Although sentences occur in which the key term or phrase lies buried deep down in the structure, preliminary observations indicate that there are many others in which the semantic hierarchy closely parallels that of the syntactic structure. This suggests that more sensitive vocabulary statistics for purposes of automatic abstracting may be obtainable by considering only words occurring in positions above a predermined cut-off level in the sentence structure. Alternatively, one might count occurrences of words on each level, and then multiply by a fixed weighting factor in each instance before taking the overall totals."[5]

---

[1] Lesk and Storm, 1962 [358], pp. I-2, I-4 ff.

[2] Ibid, p. I-31.

[3] National Science Foundation's CR&D Report No. 11, [430], p. 21.

[4] Jacobson, 1963 [292], p. 191-192.

[5] Plath, 1962 [474], p. 190.

3.4 Quality of Modified Derivative Indexing by Machine

Most of the modified derivative indexing techniques that have been proposed to date have few or no indexing results to provide comparative data for purposes of evaluation. Moreover, those techniques which are primarily directed to the generation of document abstracts rather than indexing terms have been reported to date with a paucity of actual examples. [1] One of the main reasons for this lack of product-effectiveness data is unquestionably the high cost and difficulty of obtaining substantial corpora of representative document text in machine-readable form. For the most part, the few examples of automatic abstracts produced by machine are sadly lacking in pertinency, relevancy, [2] and in continuity for scanning or reading by comparison with conventional human abstracts, whether prepared by author, editor, volunteer specialist in the subject field, or professional documentalist.

A few studies have been made for a somewhat larger numbers of examples of "auto-abstracts" with respect to differences between several different machine-extraction formulas, random sentence selections, and sentences extracted manually. A project conducted by IBM's Advanced Systems Development Division for the ACSI-matic program, (1960 [289], 1961 [290]), involved 70 to 90 articles on military intelligence items. The comparisons were of "auto-abstracts" as against titles, full texts, "pseudo-auto-abstracts" comprised of the first and last 5 percent of the sentences of each text, and sets of sentences selected randomly, without reference to conventional types of manually prepared abstracts and without respect to the quality as such. Similarly, Thompson Ramo Wooldridge data (1963 [601]) on machine-extracted and randomly-extracted, sentence sets compare these "abstracts" against manual selection of 25 percent of the sentences of each item, rather than against a conventional type of abstract.

There are however, almost no data available on the possible results of using sentence and word-group extracting techniques, applied to machine-usable texts, to the development of indexing entries rather than to the generation of substitutes for document abstracts. For this reason, as well as because discussion of the difficulties of evaluation in general will be deferred to a later section of this report, the question of the quality of modified derivate indexing will be briefly considered below, largely in terms of non-quantitative judgments.

First and foremost, as has been noted previously, is the objection that word-indexing typically produces redundancy, scatter of references among synonyms and near-synonyms, inclusion of many irrelevant entries at high page and user-scanning costs, omission of many properly indexable topics or points of interest because the authors did not emphasize them or used new and unusual terminology to describe them, failures to achieve consistency both of reference and index-vocabulary control for the papers of more than one author, and the like.

---

[1] Purto expresses regret that the studies of Agrayev and Borodin, intercomparing results of human abstracting, use of Luhn's method, and their own modification, used only a single paper (1962 [484]). Storm, (1961 [577]), evaluating the initial noun occurrence technique as a measure of sentence and index-term extraction significance, reports results for only two papers, both by Quine. Only nine articles, with no more than 40,000 words of text in toto, were used by Edmundson, Oswald and Wyllys in their 1960 experiments ([180]).

[2] Compare, for example Lesk and Storm, 1961 [358], pp. I-29 and I-30 as follows: "A final problem is the ambiguity that may arise by removing two sentences from context; two sentences alone do not always permit comprehension. Worse yet, the meaning may actually be inverted upon removal from context. For example... a quote is selected which an unsuspecting reader might think the author supports, when he is really attacking the position."

Additional difficulties are engendered, for word indexing by machine from text as against word indexing by people, because of complexities required in programming to achieve recognition of even such simple indicia as endings of sentences, [1/] inconsistencies of capitalization, [2/] and misspellings. [3/] Context distinctions between multiple meanings of homographic words are even more difficult. Difficulties in achieving good indexing quality are increased if only titles are used; those of keystroking and machine cost requirements increase as the amount of input material grows.

For these reasons, early criticisms such as those of Bar-Hillel are largely as pertinent today as they were when statistical techniques for computer generation of document extracts and index terms were first proposed. For example:

> "There can be no doubt but that computers are in a position to select out of the words or word-strings occurring in the encoded form of the original document those words or strings which fulfill certain formal, statistical conditions, such as occurring more than five times, occurring with a relative frequency at least double the relative frequency in general... However, it is ... unlikely that the set obtained thereby will be of a quality commensurate with that obtained by a competent indexer. First, there will be serious difficulties as to what is to be regarded as instances of the same word ... Second, there arises .. the problem of synonyms. Third, and most important, this procedure will yield at its best a set of words and word strings exclusively taken from the document itself." [4/]

On the other hand, there are many situations where, because of time factors or lack of conventional indexing resources, even unmodified derivative indexing by machine is itself of value and therefore modifications to improve the quality of results, whether made by man or by machine, may be well worthwhile. As Anzlowar claims: "The increasingly widespread KWIC indexes ... can save so much in time and effort that they surely deserve better than the somewhat haphazard 'slash-dash -ing' now done in most in most instances as the only cerebral operations thereon." [5/]

[1/] See Luhn, 1959 [384], p.22: "Amongst the difficulties encountered in the processing of machine readable texts, inconsistencies in the use of punctuation marks, compounds, capitals, spacing and indentations have been a problem way out of proportion with respect to the simple functions these devices stand for. For instance, even with the aid of a dozen different tests performed by the machine, the true end of a sentence cannot be determined with certainty."

[2/] See Artandi, 1963 [20], pp. 52ff, on problems of capitalization of proper names.

[3/] See Wyllys, 1963 [653], p. 15.

[4/] Bar-Hillel, 1962 [35], pp.417-418.

[5/] Anzlowar, 1963 [16], p. 104.

Modifications to derivative indexing techniques that tend toward normalizations of terminology and word usage, and increasingly sophisticated proposals for machine use of syntactic, semantic, and contextual clues hold out the promise of transition to more truly "subject" indexing and to automatic assignment indexing systems.

4. AUTOMATIC ASSIGNMENT INDEXING TECHNIQUES

Answers to the question of whether indexing by machine is possible are actually dependent in part on how the question of whether what can be achieved by machine is or is not properly termed "indexing" is answered. If "indexing" is defined as being more than the mere extraction of words from titles, abstracts, or text, then automatic derivative indexing, even when augmented by various modifications, normalizations, and editings, does not provide affirmative evidence. In the case of concept-oriented definitions of indexing, the question becomes one of whether or not automatic assignment indexing is possible. Experimental evidence suggesting that it is will be presented in this section.

We should note first, however, that just as there are differences of opinion as to what "indexing" means so there are similar differences, with respect to whether or not it represents concepts rather than extracted words. There are also a number of conflicting definitions of what is meant by "indexing" in contradistinction to "classifying". For some, the latter difference is related to questions of the number of labels or surrogates assigned to a single item to represent its subject contents, ranging from the assignment of a single subject category in a classification scheme involving mutually exclusive classes to the assignment of a number of terms or descriptor each standing for one of a number of aspects of the subject. For our purposes, however, we shall regard both the case of indexing with a number of descriptors and that of classifying to a single category or subject heading as being within the province of automatic assignment indexing, reserving the term "automatic classification" for the case where the machine is used to establish the classification or categorization scheme itself.

Actual experiments in automatic assignment indexing by Borko, Borko and Bernick, Maron, Salton, Stevens and Urban, Swanson, and Williams will be discussed briefly below. These discussions are generally in chronological order with respect to first reporting of results, except that the Salton-Lesk-Storm work reflects a somewhat different principle of assignment from the methods using clue word approaches and it is therefore described after these others have been discussed. Some of the similarities and differences between the various methods are then indicated. A brief final subsection covers related assignment indexing proposals for which experimental data is not available or has not as yet been reported in the literature.

4.1 Swanson and Later Work at Thompson Ramo-Wooldridge

Research on fully automatic indexing as well as on full text searching and retrieval at the Ramo-Wooldridge Corporation has been reported as being under way at least as early as the spring of 1958. [1] As described elsewhere in this report, experiments in search and retrieval based upon full natural language text had used as test items short articles in the field of nuclear physics. In additional experiments representing a preliminary "clue word" approach to possibilities for automatic indexing procedures, some of this same material was used.

---

[1] National Science Foundation's CR&D rept. no. 2, [430], p. 32.

In these additional experiments, 27 articles in the nuclear physics subject area were included in a corpus of 100 articles, the remainder covering a variety of topics. Frequency counts of word occurrences for the physics material were obtained and the 12 most frequent words that were judged to be discriminatory for the subject were selected. The hypothesis was then tested, that if any document pertained to nuclear physics it would contain at least two of these words. Retrieval was achieved for 25 of the 27 documents and the two "irrelevant" documents also retrieved did include information at least peripherally related to the subject. It was thus evident that the retrieval effectiveness of automatic recognition of nuclear physics subject material in the general collection was considerably greater than the average effectiveness of retrieving responses to the highly specific search questions in nuclear physics that had been used in the full text searching experiments (Swanson, 1961 [586]).

This second set of experiments provided a transition from the full text searching work, which if it can be considered indexing at all is obviously derivative indexing, to work in the application of an automatic assignment indexing method to 1,200 newspaper clippings (Swanson, 1962 [584], 1963 [580]). These were brief news items for which machine-readable texts in the form of punched paper tape were available. Thesaurus-groups of words likely to be associated with each of 20 to 24 subject headings were first compiled on the basis of human analysis of 1,000 or more representative items. These word groups were further screened so that no word appeared in more than one group and so that each word retained should be uniquely indicative of the particular subject category. In the machine assignment procedure, subsequently, if a word occurs that belongs to a particular thesaurus group, the corresponding subject heading is assigned to the item in which that word occurs.

Results achieved with this technique appear to be highly promising, at least for this type of material. Swanson reports as follows:

> "Approximately 1,200 brief news items were classified into 20 nonhierarchical subject categories, both by a human and a machine procedure. Each item was assigned on the average to about four categories. The results of the two processes were compared. With the human process as a standard, the machine missed only seven percent of the correct subject assignments and made a number of irrelevant assignments equal to about 17 percent of the total. Nearly 40 percent of the automatic subject assignments judged finally to be correct were missed by the human catalogers."[1]

While this accomplishment is actually due to the extensive human effort to compiling, organizing, and pruning of the uniquely indivative word lists, it is pointed out that this intellectual effort and the programming tasks need to be done only "once and for all".[2] It is further pointed out that garbles or misspellings in the input text do not appear to affect the procedure, there being enough redundancy in the messages so that even if one or two clue words are missed, others will be present.[3]

---

[1] Swanson, 1962 [584], p. 468.

[2] Ibid, p. 469.

[3] Swanson, 1963 [580], p. 5.

Swanson and his TRW associates have further proposed extensions of the prespecified unique clue-word technique. For example, it is suggested that machine processes of comparing words of titles, subtitles and chapter headings to lists of possible subject heading can be extended in sophistication by machine lookups of synonym groups and of characteristic subject-word associations. [1/] Frequency weightings may be taken into account, and similar measures of association and subject-indicativeness may be developed for phrases as well as for individual words. [2/] In general, however, the apparent success of this clue-word technique in tests to date should be considered in the light of the special character of the items, their extreme brevity, and the high probability that the fact-word incidence involved in news reporting is not typical of less popular and less factually oriented materials. [3/]

Continuing work along similar lines has been carried forward at Ramo-Wooldridge in the "Word Correlation and Automatic Indexing Program" sponsored by the Council on Library Resources (1959 [490] and [491]). Here, the objectives are to develop and apply clue-word techniques to material that is much more representative of the scientific and technical literature. The thesaurus-groups, now called "indexonym" groups, are made up of words and phrases selected by extensive human analysis as being significantly "useful-for-retrieval-purposes".

New items would be processed in a word and phrase lookup operation, with each word or phrase being initially assigned the identifier number codes of all groups to which it belongs. However, unless a particular group's number is repeated several times within the space of a few paragraphs, it is not used as the basis for the actual assignment of an index tag. Provision would be made for calling human attention to items having a number of words that are not deleted by processing against a "useless-for-retrieval purposes" list, but that are not found in any of "accepted" groups. It is suggested that in this way it should be possible to "ascribe measures of automatically recognizable 'newness' to technical articles". [4/]

4.2 Maron's Automatic Indexing Experiments

By April of 1959, the reports of work at Thompson Ramo-Wooldridge on automatic indexing and related problems submitted for the Current Research and Development in Scientific Documentation series included reference to Maron and a "probabilistic model for the assignment of index tags", as well as to Swanson's continuing projects. [5/]

[1/] Swanson, 1962 [584], p. 469.

[2/] Swanson, 1963 [580], pp. 1-2.

[3/] See also Mooers, 1963 [424].

[4/] Thompson Ramo Wooldridge, 1959 [491], p. 2A.

[5/] National Science Foundation's CR&D report No. 5 [430], p. 34.

In addition to his work on probabilistic indexing with emphasis on relevance weightings for index tags manually assigned, Maron has actively explored automatic assignment indexing techniques. The approach is also probabilistic, with emphasis on the statistics of association between content-indicative clue words and subject headings manually assigned to sample documents. The experimental corpus consisted of a group of abstracts in the field of computer technology indexed to 32 subject categories designed for the purposes of these investigations.

Common words such as articles and prepositions were first excluded. Next, words occurring less than three times were purged and words such as "data" and "computer" were also rejected because they occur so frequently in this literature. Approximately 1,000 words remained after these purging operations. After sorting the source documents to their most appropriate subject categories, statistical frequencies were obtained for the co-occurrences of the candidate clue-words with the categories and the resulting listings were manually examined to determine which words peaked in a particular category. Eventually, 90 such words were selected.

The occurrence of one or more of the 90 clue-words in the text of new documents was then used to predict the subject category to which the new item should belong. [1] Tests were run with two groups of documents, one consisting of the source items from which the statistical frequency and word list data had been obtained, and the second group consisting of 145 genuinely new items. For the latter group, twenty documents contained no clue words whatever and forty items had only one. For the remaining 85 items having two or more clue words, the results of the computer assignment program were predictions of the correct category in 44, or 51.8 percent, of the cases. [2] Results using the source documents were significantly better, as expected, with 84.6 percent accuracy of category prediction for 247 items. Results were also related to the number of clue words that occurred in the test items, with a prediction accuracy of only 48.7 percent for items with a single clue word rising to 100 percent probability of correct assignment if six or more clue words occurred.

Trachtenberg (1963 [608]) has also considered a probabilistic approach to automatic indexing and categorization of documents, similar to that of Maron. He suggests the investigation of two information theoretic measures with reference to determination of which of various possible clue words are significantly discriminating with respect to the different categories. He further suggests experiments using 90 clue words and the corpus used by both Maron and Borko, but no actual results have as yet been reported.

4.3 Automatic Indexing Investigations of Borko and Bernick

At the System Development Corporation, the work of Borko (1960 [73]), and of Borko and Bernick (1962 [77], 1963 [78], 1964 [79]) in the area of automatic indexing has involved both automatic assignment indexing and automatic classification techniques. They have not only reported actual indexing results but have provided data for the intercomparison of their techniques with the experiments of Maron for the same source material.

---

[1] Note that the word itself is not necessarily used as an index tag or label, as is the case for derivative indexing using an inclusion list approach. This is an important distinction.

[2] Maron, 1961 [395], p. 257.

The original Borko approach was based on the principles of factor analysis as these had been developed for the analysis of multivariate date, especially in the field of psychology. Borko's first experiments were directed to a corpus consisting of 618 abstracts in the field of psychology, amounting to approximately 50,000 words of total text and 6,800 different words. These words were sorted by computer program into an order reflecting their respective frequencies of occurrence. For the approximately 200 words that occurred twenty or more times in this corpus, the investigator himself selected 90 words to serve as index (or, better, index-clue) terms. A matrix was then developed for the frequencies of co-occurrence of these words and the documents in which they appeared. From this, a 90 x 90 correlation matrix was computed as follows:

> "To compute the correlation coefficient . . . we used the following formula
>
> $$r_{xy} = \frac{N\Sigma xy - (\Sigma x)(\Sigma y)}{\sqrt{[N\Sigma x^2 - (\Sigma x)^2]\,[N\Sigma y^2 - (\Sigma y)^2]}}$$
>
> Where N is equal to the number of documents (618) and x and y are the terms being correlated." [1]

The term-correlation matrix was then factor analyzed and the first ten eigenvectors were selected as factors to be rotated and interpreted. Borko emphasizes that:

> "The interpretation must be made by the investigator and is based upon his knowledge of the analytic procedures and the subject matter. There is, therefore, a degree of subjectivity in the names selected for each factor. These names may be regarded as hypotheses about the factor meaning." [2]

Following the derivation of these "classification categories" by means of the factor analysis technique, new items may be assigned to the categories on the basis of words occurring in their texts (abstracts) in accordance with the following procedural steps:

> "1. Each document, in machine readable form, is analyzed by the computer. A list of the index terms and their frequencies of occurrence in each document is recorded.
>
> "2. The category or categories containing the index term is assigned a value equal to the product of the number of occurrences of the word in the abstract and the normalized factor loading of the word in the category. If more than one index term appears in a category, the products are summed.
>
> "3. After each index term has been considered, the category having the highest numerical value is selected." [3]

---

[1] Borko, 1961 [73], p. 283.

[2] Ibid, pp. 285-286.

[3] Borko and Bernick, 1962 [77], pp. 7-8.

The choice of 90 clue words in Borko's work with abstracts in the field of psychological literature was apparently dictated by a matrix size which would be convenient for computer manipulation. [1] However, it happened to coincide with the number of clue words used by Maron in his experiments. Advantage was taken of this coincidence to obtain comparative data on the performance of the two assignment-indexing techniques as applied to the same material. The 260 computer literature abstracts used by Maron, as source documents were processed to derive a correlation matrix for Maron's 90 manually selected words, which was then factor analyzed. Several sets of factors were extracted, rotated, and the results studied, with a final selection of 21 categories.

Since these automatically derived categories did not coincide with Maron's original 32, it was necessary to analyze manually the total group of 405 abstracts (260 "source" and 145 "test" items) and assign them to the new categories, then to study the documents falling into each factor-analytically derived category to determine which of Maron's 90 clue words were category-indicative, and finally to substitute these words in the Bayesian equation used by Maron so as to predict which of these classification categories his probabilistic method should obtain.

The same two sets of 260 "source" and 145 "new" abstracts used by Maron were then submitted to the computer assignment program which compares the clue words of a new item with the numeric values of the predictor words for each factor category, then computes the score for each item in all categories, and assigns the category with the highest score to the item. For the source items, Borko and Bernick's results showed 63.4 percent correctly classified, by comparison with the 84.6 percent correctness score originally obtained for them in Maron's experiments. For the new items the factor analysis method scored 48.9 percent correct assignment by comparison with Maron's original 51.8 percent. [2] The later investigators therefore concede that the performance of Maron's technique was somewhat superior for the same items using the clue words originally selected by Maron.

Further experimentation was then carried out (Borko and Bernick, 1963 [78]) using word frequency data for the selection of a new set of 90 clue words and a classification scheme for 21 categories was again automatically derived. The 405 abstracts were again manually classified to these machine-derived categories by five subject-matter specialists and the two investigators. Comparative data were then obtained for both the Maron assignment formula and the modified classification system assignments in terms of agreement with the manual assignments.

For the source items, the percentage of machine assignments agreeing with those made by people was 62.7 when the Bayesian probability formula used by Maron was applied and 61.2 for the factor analysis score system. For the new items, the corresponding correct percentages were 57.9 and 55.9. Additional data compared the effects of using the original Maron words and the frequency-based word set (Borko's words) for the same probability formula assignment method. While there was an overlap of approximately 50 percent between Maron's words and Borko's words, the findings indicated that:

---

[1] Now increased to 150 x 150.

[2] Borko and Bernick, 1962 [72], pp. 9-10.

> "... The index words selected by Maron are decidedly specific to the documents from which they were derived and are of less generality than the frequency based terms. The Bayesian formula coupled with the Maron words correctly predicted the classification of 79.6% of the documents in Group I ['source items'] but only 43.9% of the documents in Group II ['test items']. The coupling of the Bayesian formula with the Borko words resulted in a slight decrease in the percentage of Group I documents whose classification was correctly predicted (62.7%) but increased the percentage of correct prediction for Group II documents to 58.0%." [1]

Other findings from the later experiments indicated that despite the differences in the two word-sets, the factor categories derived from them were very similar. It was also found that, at least for the source items (Group I), the two machine techniques and the manual process classified 56.1 percent of the items into the same categories. It should be noted, however, that in the case of the automatic assignment methods: "Eleven documents contained no clue words and could not be automatically classified by either system." [2]

4.6 Williams' Discriminant Analysis Method

The work of Williams in automatic assignment indexing, reported in the fall of 1963 [642], has also involved tests on abstracts of the computer literature, directly comparable to but not necessarily identical with those used by Maron and by Borko and Bernick. This work at IBM's Federal Systems Division, Bethesda is based in part on earlier work by Meadow which involved computer studies of matching functions for document word lists and category word lists for test items drawn from such fields as psychology, law, computer abstracts, and news items. [3] What has subsequently been developed is termed a "discriminant" method which begins with hierarchical classification structure of pre-established subject categories and with a small set of sample documents previously indexed by people into these categories. Frequency counts of words in each of the sample documents lead to computations, for each category, of the theoretically probable frequencies of its most statistically significant words. For new items, observed word frequencies are compared with the theoretical word-category associations and a relevance value is computed for the item in terms of each category.

The corpus selected for experimentation consisted of 400 items from "Computer Abstracts on Cards". [4] These had previously been indexed using a classification structure of 15 major categories, each of which is divided in turn into 10 subcategories. The experimental sample, however, was so selected as to provide exactly 15 "source" items and 5 "new" items for each of 5 subdivisions of 4 of these major categories.

---

[1] Borko and Bernick, 1963 [78], p. 23.

[2] Ibid, p. 11.

[3] Williams, 1963 [642], cites H. B. Meadow, "Statistical Analysis and Classification of Documents", IRAD Task No. 0353, FSD IBM, Rockville, Maryland, 1962, but this is apparently a company-confidential document, containing proprietary information. Meadow gave an informal report on her work at the Computing Center seminars, University of Maryland, in March of 1963.

[4] Available on a subscription basis from Cambridge Communications Corporation, Cambridge, Mass.

Discriminant coefficients were then computed at both the major and minor levels for all words occurring in the sample items falling into one of the 20 groups in accordance with the formula:

"The discriminant coefficient is:

$$\lambda_i = \sum_{j}^{n} \frac{(P_{ij} - \bar{P}_{ij})^2}{\bar{P}_{ij}}$$

Where:

$$P_{ij} = f_{ij} / \sum_{i}^{m}$$

The relative frequency of the ith word in the jth category.

and

$$P_{ij} = \frac{1}{n} \sum_{j}^{n} P_{ij}$$

The mean relative frequency per category of the ith word. [1]

These coefficients are used both to set up threshold values to determine which words should be used in the assignment formulas and to assign weighting factors to the words themselves.

The results of the experiments to date are based on 83 items from the "reference set" which were not used as source items. For 63 items, 78 percent were correctly classified at the level of a single major category (e.g., "Programming", "Hardware Design") and also correctly classified at a single subcategory level, (e.g., "Programming Languages", "Semiconductor Devices"). The 20 remaining items were classified to one major category with an accuracy of 95 percent and to two minor level subdivisions with accuracies of 60 percent and 75 percent. Additional investigations were made on the effects of using a discrimination threshold to eliminate insignificant words from consideration and on the use of weighting factors in the assignment calculations.

4.5 SADSACT

Stevens and Urban at the National Bureau of Standards (1963 [569, 570]) have also explored an automatic indexing technique that uses, as in the experiments of Williams, a teaching sample or reference set of previously indexed items to form patterns of word and index-term assignment associations. However, there are much less formal requirements for computing correlation coefficients and no consideration is required of either the theoretical probabilities of word occurrence by category or of discrimination coefficients and thresholds. Instead, the technique involves ad hoc statistical associations between the words occurring in the title and in the abstract of a sample item and the descriptors previously assigned to that item. A master selection-word vocabulary is thus built up where each word is listed in terms of the frequencies of its co-occurrence with each of the descriptors with which it has co-occurred, regardless of whether or not such prior associations are either revelant or significant. No attempt has as yet been made to "purge" the resulting association lists. Instead, reliance is placed on the patterns of multiple word usage and of redundancy of words used in titles and cited titles of new items to minimize the effects of irrelevant or accidental prior word-descriptor associations and to enhance the significant ones.

---

[1] Williams 1963 [642], p. 163.

The SADSACT method (for "Self Assigned Descriptors from Self and Cited Titles") proceeds with the assumption, which it shares with the arguments for citation indexing previously discussed, that the literature references cited by an author are indicative of the subject content or contents of his paper. [1] For the automatic indexing of new items, their titles and the titles of up to ten bibliographic references cited are keystroked, converted to punched cards, and fed to the computer. This input material is run against the master vocabulary to obtain for each input word which matches a vocabulary word a "descriptor-selection score" for each of the descriptors previously associated with that word. These scores are summed up for all words and at an appropriate cutting level those descriptors having the highest scores are assigned to the new item.

Preliminary results based on the titles and cited titles of items that were "source items" in the sense that their titles and abstracts had been used in the teaching sample were reported at the NATO Advanced Study Institute on Automatic Document Analysis held in Venice in July, 1963. For 30 items drawn from such subject fields as computer technology, information selection and retrieval, mathematical logic, pattern recognition, and operations research, all of which had previously been indexed by ASTIA personnel in 1960, the machine assigned 64.8 percent of the descriptors previously assigned. Subsequent tests on genuinely new items, however, resulted in a drop to only 48.2 percent "hit" accuracy.

These "new" item results were also evaluated by having several representative users of the collection analyze the test items and assign descriptors to them from a list of the descriptors available to the machine. The extent to which the descriptors assigned by machine were also independently chosen by one or more of these indexers was then checked. In general, the fewer descriptors assigned by the machine, the better was the human agreement, ranging from 47.4 percent overall in the case where the machine had assigned twelve descriptors to each item to 76% agreement where the machine assigned only one. In particular, for ten items which were analyzed by five different indexers, the chances that one or more would also select the machine's first choice (highest scoring) descriptor averaged 90 percent.

4.6 Assignment Indexing from Citation Data

Certain phases in the program of investigation of information selection and retrieval problems at the Harvard Computation Laboratory have been mentioned previously. The work of Storm and of Lesk and Storm on the use of first-noun-occurrences as selection clues for both automatic indexing and abstracting was discussed in connection with techniques for improved derivative indexing. The studies on citation indexing have included, as noted, experiments to assign indexing terms to a new document by finding the indexing terms previously assigned to the five most "related" documents, where "relatedness" is a function of the similarity in citation patterns as between the new document and items already in the collection. The results of such index term assignments are reported as identical to those made by human judgment approximately 50 percent of the time. [1/]

---

[1] If necessary or desirable, however, abstracts or portions of text can be used in addition to or in lieu of the cited titles.

More specifically, in an experiment using documents drawn from a small collection in the fields of mathematical linguistics and machine translation, a new item was compared in terms of its citation data with the citation similarity data previously determined for earlier documents, and the set of five related documents was selected using the magnitude of the row similarity coefficients obtained from links of length one and two. All index terms occurring at least twice in the set of terms assigned to these related items were then assigned to tne new items. For the ten "typical" new item cases, for which comparative data are shown, the citation data assignment method correctly assigned, on average, 47.6 percent of the terms assigned manually to the same items. [2/]

A slightly more sophisticated indexing term assignment formula, described by Lesk, was applied to additional test cases, but "failed to raise accuracy above fifty percent". [3/] For five typical new cases, the improved method correctly assigned 11 of the 20 terms manually assigned to these items, or an average accuracy of 55.5 percent. [4/]

4.7 Similarities and Distinctions among Assignment Indexing Experiments.

In Table 2 some of the key points of the various automatic assignment indexing experiments we have discussed above are summarized. Certain similarities, distinctions, and differences are to be noted. Borko and Bernick use the same corpus as did Maron and also re-apply Maron's formula to a different clue-word set for the same material. Williams uses material similar to the Maron-Borko computer corpus. The SADSACT tests also use some items that might be included in the Maron-Borko and Williams corpora. The Swanson experiments with newspaper clippings represent a quite different class of material consisting of brief, terse, factual messages.

[1/] Lesk, 1963 [357], p. V-8.

[2/] Salton, 1962 [520], p. III-41, Table 9.

[3/] Lesk 1963 [357], p. V-7.

[4/] Ibid, p. V-8, Table 3.

Table 2. Summary of Automatic Assignment Indexing Test Evaluations

Investigator	Principles and Methods	Materials Used	Tests	Remarks
Maron	Statistical probabilities of association between clue words and pre-established subject categories. Source items manually indexed to 32 categories. A subclass of words occurring in the corpus selected as clue words, and statistical correlations obtained for 90 such words with categories assigned. Correlation data and Bayesian probabilities used to assign categories to new items.	Corpus of 405 items selected from computer abstracts, PGEC, 1959. Full text, 20, 000 words of which 3, 263 were different words.	For 260 source items, 12 did not contain any clue words, 247 were indexed, 1 contained an error preventing processing. For the 247 source items indexed, probability of top-ranked category being correct = 84. 6%. For 145 new items, 20 not indexed because they contained no clue words. In 85 cases where at least 2 clue words occurred, probability of correct category assignment = 51. 8%.	Considerable manual inspection and judgment involved in the selection of clue words. Some new items cannot be processed, because they contain no clue words.
Borko	Factor analysis to determine distinctive grouping of clue words. Word frequency counts made, 90 of the 2. 0 most frequent non-common words manually selected. Correlation matrix computed, factors rotated and interpreted.	Psychological abstracts. 618 abstracts, 50, 000 text words; 6, 800 different words.	Factors selected were judged to be compatible with but not identical to subject classification terms used for these items by the American Psychological Association.	Some new items cannot be processed, because they contain no clue words.

Table 2 (cont.)

Investigator	Principles and Methods	Materials Used	Tests	Remarks
Borko and Bernick	Factor analysis to determine distinctive groupings of clue words. Maron's 90 clue words used for word-word correlation and factor analysis. 21 factors developed, and items manually re-indexed to these categories.	Same corpus as Maron, 405 computer abstracts, of which 260 used to establish factors, 145 as new items.	Detailed comparison with Maron's technique. For the source items, 63.4% were correctly classified. For the new items, 46.5% correctly indexed, and 48.9% were correct for those items in which 2 or more clue words occurred.	Some items cannot be processed because they contain no clue words.
Swanson	Text word lookup against clue word lists, constructed by careful analysis of sample items to be exclusively indicative of a particular subject heading. Machine assigns a subject heading to an item if any word on its list occurs in that item.	Brief news dispatches available on teletype tape, wide diversity of topics. From study of several 1,000 items, 24 subject headings established and word lists selected, averaging approximately one hundred per category. 775 new items then tested.	Machine assignments compared to manual subject indexing. For a first batch of 500 items, 569 assignments of correct headings, 119 assignments of irrelevant headings, and 32 correct headings missed. The clue word thesaurus was then revised. For 275 additional test items, results showed 282 correct assignments, 29 irrelevant assignments, 1 missed. For total, averages of 17% irrelevant assignments, 3% missed. For 200 items, machine and manual assignments were compared with respect to 5 of the subject categories, with the following results: Man Machine Irrelevant 4 25 missed 46 4 correct 75 116	

Table 2 (cont.)

Investigator	Principles and Methods	Materials Used	Tests	Remarks
Stevens and Urban	Teaching sample for machine compilation of co-occurrence data for words in titles and abstracts with descriptors assigned to these items. Words in titles and cited titles of new items then run against master list of previous word-descriptor association to derive descriptor-selection scores, highest scoring descriptors (e.g., up to 12) assigned. Associations derived for 1,600 words co-occurring with any of 70 descriptors previously assigned.	Two teaching samples, approximately 100 items each with 70% overlap, drawn from items indexed by ASTIA. For new items titles and up to 10 cited titles.	For 59 test items, assignments of descriptors that had occurred for at least 3% of the sample items agreed with ASTIA assignments 58.1%. However, for all descriptors assigned by ASTIA, many not available to machine, overall machine accuracy = 40.1%. For 20 items, independently evaluated by several typical users, the chances that one or more people would agree with the machine assignments ranged from 47.1% when 12 descriptors were assigned to 75.0% average agreement with the machine's first choice.	All test items could be processed and up to 12 different descriptors assigned to each, but some descriptors used in manual indexing of these items are not available to the machine.
Williams	Discriminant analysis. Sample items previously indexed to a 2-level classification system were subjected to word frequency counts and the theoretical frequencies of the most significant words in each category were compiled. For new items, observed word frequencies' compared with theoretical frequencies for each category, highest scoring assigned.	Items from "Computer Abstracts on Cards" indexed to 15 major categories each divided into 10 minor categories. 300 abstracts selected to provide equal distribution to 20 sub-categories, 5 each in 4 major categories. Additional items for test similarly selected.	For 63 new items assigned by machine to 1 major and 1 minor category, 78% correct at major level, 64% correct at minor level. For 20 items classified to 1 major and 2 minor categories, 95% correct at major level, 60% and 75% correct at the minor level.	

None of the experiments has so far encompassed testing of anything but very small test item samples and the dangers of extrapolating from so small and so specialized bodies of data should be clearly recognized. Mooers identifies these dangers in terms of

"The Silent Postulate:

That (real people) (real documents) (real jobs to do) can somehow

be eliminated from the experimental study, and that (role-playing people) (substitute documents) (imaginery jobs)

can be substituted and still give valid experimental results." 1/

In most of the experiments in automatic indexing conducted to date, indexing and classification schedules have been especially designed, or evaluations made, specifically for the purposes of these tests. Williams, however, stresses the point that the material used in his experiments had been "classified by professional indexers for the purposes of actual retrieval." 2/ A similar claim can be made for SADSACT, as noted by Mooers. 3/ Swanson's news item work also obviously relates to real items and implies a real job to be done, but is directed, as noted, to a class of material not generally comparable to that found in documentation operations on scientific and technical literature.

In contrast with the treatment of each document as a self-contained entity without reference to any other documents, as is the case for derivative indexing, all of the automatic assignment indexing experiments, by virtue of the fact that they are assignment techniques, do to some extent embody the effects of a consensus of a particular collection, or a consensus of prior indexing, or a consensus of human subject content analysis applied to sample documents, or some combination of these effects. The SADSACT method, in addition, wherever cited titles are available for new items, takes advantage of terminology other than the author's own as a source of clue words. Other proposed methods of assignment indexing, such as the use by Salton, Lesk, and Storm of citation-pattern similarity data, would carry the latter principle even further.

1/ Mooers, 1963 [424], p. 5.

2/ Williams, 1963 [642], p. 162.

3/ Ibid, p. 5.

4.8 Other Assignment Indexing Proposals

A few additional automatic assignment indexing proposals are under development. Examples for which experimental data is not as yet generally available include, for example, work at EURATOM, some preliminary experiments at Chemical Abstracts Service, work at General Electric, Bethesda, the proposed "Multilindex" system of Information Systems, Inc., investigations by Slamecka and Zunde, and a special purpose development project at Goodyear Aerospace.

Meyer-Uhlenried and Lustig report for the EURATOM developments as follows:

> "...Procedures are being developed which allow based upon given keyword lists first for abstracts: (a) to assign significant keywords and (b) based upon hierarchically organized keyword lists, to assign the documents in question to specific subject fields.
>
> "Experiments were made at first on narrow fields with so-called micro-thesauri, they showed encouraging results when automatic and manual assignment were compared. Positive results depend of course on the quality of the abstracts and the significance of the words employed in them. It remains to see how far this favorable prognosis is confirmed by keyword collections of more complex contents." 1/

Friedman and Dyson (1961 [203]) have reported on manual experiments designed to relate words occurring in a sample of abstracts from a particular section of Chemical Abstracts to the title or heading for that section. Significant words in these abstracts were counted and the number of occurrences as well as the number of different abstracts in which they appeared were determined, with a rank order listing as a result. It appeared, from inspection, that it should be feasible to develop, for each CA section, a relatively small vocabulary of words that would be descriptive, and indicative of, the subject matter contained in it. They conclude: "In our opinion, the results were significant, the small vocabulary of words did select a large percentage of the abstracts in the section it was based on." 2/

A project at Information Systems Operations, General Electric, on possibilities for automatic indexing and abstracting of text has been reported in the November 1962 issue of Current Research and Development. 3/ The META project (Methods of Extracting Text Automatically) is said to be concerned with the use of statistical, linguistic, and semantic criteria for analysis and selection of significant words and significant sentences from text. Computer programs are being developed in modular fashion for the GE-225 computer.

1/ Meyer-Uhlenried and Lustig, 1963 [417], p. 229.

2/ Friedman and Dyson, 1961 [203], p. 10.

3/ National Science Foundation's CR&D report, No. 11 [430], p. 97.

The proposed "Multilindex" system is also based on micro-thesauri or small vocabularies designed, by human analysis, for clue-indications to a relatively narrow subject field, together with potential syntactic-semantic role indications built into the dictionary, again by extensive human analysis, following the approaches previously taken by A. L. (Lukjanow) Loewenthal in her suggestions for solutions to problems of mechanized translation. An unpublished proposal-type brochure describing the system was available as of December 1963.[1] As of that date, also, demonstration printouts were available from an IBM 1401 Fortran program, illustrating an index compiled from abstract-text input and a 1,200-word dictionary for documents in the field of space antenna tracking radar.[2] A repetoire of 350 "concepts" or indexing terms was involved, with an average of 10 assigned to 22 test documents, many of these assigned terms being identical to words occurring in either the title or the text of the abstract of the item.

Slamecka and Zunde have investigated the extent to which the "notations-of-content" in the system developed by Documentation, Inc. for NASA's STAR might be derived by machine techniques from the text of the abstracts with enough normalization-standardization via inclusion dictionary lookup to qualify as an assignment indexing technique. These workers claim:

> "This preliminary investigation indicates the possibility of using the computer to index documents adequately for machine retrieval by matching their abstracts against an authoritative subject-heading authority ... The inconsistency inherent in human indexing can be eliminated as the number of terms derived from any one abstract will always be the same. The abstract and its automatically derived set of index terms will always be equivalent..."[3]

A final example of other approaches to automatic assignment indexing research, not yet reported in the open literature, is an NIH sponsored project at Goodyear Aerospace, in cooperation with the Universities of Minnesota and Rochester and Western Reserve University, looking toward an automatic classification procedure based on word coocurrences for a set consiting of 100 four-to-five page documents in the field of diabetes literature. Programs for statistical analyses of the full text of these documents, all of which have previously been processed for the manual W. R. U. "telegraphic" abstracting system, are being developed.[4]

5. AUTOMATIC CLASSIFICATION AND CATEGORIZATION

In all the experimental work, to date, that has been directed toward the use of computers and other machine-like techniques for the automatic indexing of documents, a dichotomy can be observed. There is, on the one hand, a spate of examples of automatic derivative indexing where words used by the author himself or by human analysis are sorted and arranged, by machine, to provide index listings, announcement bulletins, and current awareness distribution notices. There are also, on the other hand, at least a few instances of investigations where the machine assigns category labels, indexing terms, or "heads" and "headings" from a classification schedule, to new items.

---

[1] "Description of MULTILINDEX. A mechanized system for indexing documents, storing information, retrieving information", P.S. Shane, Dec. 4, 1963, Information Systems, Inc., 7720 Wisconsin Avenue, Bethesda, Maryland.

[2] Private communications, A.L. Loewenthal and P.S. Shane, Dec. 11, 1963.

[3] Slamecka and Zunde, 1963, [561], pp. 139-140.

[4] E. Tuttle, private communication, Oct. 30, 1963.

In general, as Needham [1] points out, proposed automatic assignment indexing procedures can be investigated with reference to a previously existing index term vocabulary, an existing classification system or schedule, or to specially designed vocabularies and subject heading lists. On the other hand, if it is not known how well existing systems do in fact characterize documents and if it is not known whether all pertinent properties of the documents have been consistently identified, then it may be preferable to develop methods for assigning documents to the appropriate class in a classification system which is itself set up automatically. [2] Needham also suggests still a third possibility: that of setting up automatically a classification within which the subsequent classifying of documents is done by hand.

The principal experimental results, to date, of attempts to achieve automatic classification of documentary items, especially in the sense of machine-generated groupings or categorizations of such items, have been those of applying techniques of "clumping", [3] factor analysis, and "latent class analysis". [4] We shall briefly consider below some typical investigations into automatic classification or categorization procedures that have already had, or may have, applicability in automatic indexing techniques.

In the late 1950's, Tanimoto undertook theoretical studies of mathematical approaches to problems of classification and prediction with special reference to matrix manipulations of sets of attributes of items to be classified. [5] He also investigated theoretical aspects of automatic indexing and sentence extraction involving co-occurrences of words. While Tanimoto's studies with respect to linguistic information processing for classification purposes have apparently been limited to the theoretical considerations, similar concepts of probabilistic, computational, and matrix manipulative operations to derive and use coefficients of correlation of associations between such attributes as words occurring in text or the index terms assigned to documents are involved in the factor analysis and theory of clumps techniques as applied in actual experiments in documentary classification.

---

[1] Needham, 1963, [432], p. 1.

[2] Ibid, p. 1-2: "If we are to assign a document to a class automatically, we must have a) a list of facts about the classes which will make ascription possible: b) an algorithm, usually some sort of matching algorithm, to tell us which class best suits a document. Given a classification like the U. D. C. it is not at all obvious that a) and b) exist, or even, if they can be found. a) and b) imply a degree of uniformity about the classification which may just not be there."

[3] That is, the clustering of objects that are in some sense similar because they share certain attributes or properties, even if, and especially when, the identity of cluster-producing common properties is not known in advance.

[4] Compare Doyle, 1963 [162], p. 13; "There are other statistical techniques besides factor analysis whose output is document clusters, such as latent class analysis and clump theory, and there is a surprising increase in research in this kind of analysis just within the last two years."

[5] Tanimoto, 1958 [593], 1961 [594]. See also Borko, 1963 [76], pp. 4-5: "In 1958, Tanimoto published a theoretical paper on the applications of mathematics to the problems of classification and prediction. Specifically, he pointed out how the problems of classification can be formulated in terms of sets of attributes and manipulated as matrix functions."

5.1 Factor Analysis

The factor analysis technique which seeks to derive from word associations in representative documents an automatically generated classification schedule for use in actual indexing experiments has previously been mentioned. [1] Reasons suggested for its use in research at SDC have been reported as follows:

> "The development of automatic procedures for purposes of classification and abstracting requires the identification and specification of attributes of words or passages so that the relevancy of topics or content can be determined. Automatic procedures to detect such attributes may be based on a number of characteristics of the text: word frequencies, syntactical information, semantic information and pragmatic contextual clues. Currently, word frequency information can be generated and manipulated by automatic procedures, whereas the other attributes are not as readily handled this way. However, a correlation matrix of content words becomes very unwieldy because of its size and the complexity of relationships. For this reason, factor analysis is used to identify clusters of relationships. Current work concentrates primarily on determining the usefulness of factors identified in this way as classification and indexing schemes." [2]

As noted above, Borko and Bernick (1961 [73], 1962 [77], 1963 [78]) have applied this technique to abstracts drawn from psychological literature and to the same computer literature abstracts as had been used by Maron, (1961 [395]). This technique had also been investigated in the studies looking toward information retrieval classification and grouping undertaken at the Cambridge Language Research Unit from about 1957 onward. However, certain apparent limitations of the factor analysis approach led Parker-Rhodes and Needham to the alternative of the "theory of clumps" (1960 [465], 1961 [435, 464]). Parker-Rhodes gives the rationale, and some of the distinctions between the two techniques, as follows:

> "It has been assumed that statistical methods could be applied to the data in such a way as to reveal any objectively existing classes which may be there. The general

---

[1] Pp. 94-97 of this report.

[2] System Development Corporation, 1962 [590], p. 15.

names for the techniques evolved in this way in factor analysis. Insofar as it is practically applicable this technique has worked well enough; but...it has two limitations (a) that some classification problems are outside its scope, and (b) that it is not susceptible (at least as hitherto conceived) of adaptation computationally to the study of really large universes..." 1/

"...The procedure of factor analysis first finds certain clumps, but then, as output, it gives us vectors relating the descriptors of the universe to the clumps found...

"In most cases, factor analysis is used (especially in psychology) to delimit the descriptor space; more conventionally put, to eliminate those tests (descriptors) which have an equivocal membership in several factors (Clumps) in favor of those which, having more definite allegiances, convey more information of the kind which the analysis suggests as valuable. It is thus only related to the classification of the universe at one remove; the classification it suggests is a simple categorical classification defined by the descriptors suggested as the most valuable...

"The descriptive array of a universe is a table giving the applicability or inapplicability of each descriptor to each element. To classify the elements of the universe, we calculate for every pair of elements a similarity as a function of the corresponding rows of the descriptive array, and then regard the similarity matrix as a sufficient description of the universe. In factor analysis, on the contrary, we start with the matrix of correlations between the descriptors, each being a function of a pair of columns of the descriptive array..." 2/

Other investigators who have considered factor analysis techniques for possible applications to automatic indexing, automatic categorization of items in a collection of items, or search prescription manipulation in a mechanized selection and retrieval system include Stiles (1962 [575]), Doyle (1963 [166]), and Hammond (1962 [251]).

Stiles, whose principal experimental results relate rather to the use of statistical associations between terms manually assigned to documents for search prescription formulation and manipulation than to automatic indexing procedures as such, 3/ has also considered both automatic indexing and automatic classification approaches. Specifically, he has made at least preliminary investigations of the factor analysis technique independently developed for similar purposes by Borko. For a large collection of 100,000 items, the statistics of co-occurrence of indexing terms were in some cases not as precise as desired because the same terms were used in different senses for different items in the collection.

1/ Note that Borko himself confirms this limitation as recently as November 1963, in stating, of the CLRU work on clumps: "However, even now these techniques have been applied to a 700x700 matrix which is beyond the capabilities of presently available factor analysis programmes." (1963 [76], p. 8).

2/ Parker-Rhodes, 1961, [466], pp. 3-4.

3/ This principal concern is discussed below with reference to potentially related research, pp. 119-122 of this report.

The possibilities of using factor analysis to sort out the different meanings were therefore explored. 1/ Using an IBM 704 program, the centroid method of factor analysis was applied to a matrix of correlation coefficients of terms that had co-occurred significantly with the term "exposure". Three factors were derived, one generally relating to the corrosive effects of exposure, another to "exposure" in the sense of photographic exposure, and the third dealing with both exposure-to-weather and exposure-to-radiation. Although the results were considered quite satisfactory, more extensive experimentation and use did not appear feasible because of computer matrix manipulation limitations.

Doyle notes, in particular, that factor analysis might be used to give well-defined clusters separated one from another by clear boundaries rather than the less precise clusters found by most document grouping techniques. He emphasizes, however, that "its success in doing so of course, depends on the well-defined clusters actually being present in the data". 2/ He suggests that a combination of factor analysis and human editing to select items most typical of statistically derived categories could be valuable in such applications as the sorting of Congressional mail or the identification of trends in political or military intelligence materials free from the personal biases of an analyst.

Hammond and his Datatrol associates who have worked on an application of the Stiles association factor technique for search question negotiation to legal literature have also considered the potentialities of factor analysis. Thus they report:

> "... The present association factor gives the relationship of one term to another. A factor analysis study would allow us to determine the relationship of a single term to a group of terms. From this we could learn how terms cluster when related to the same concept." 3/

5.2 The Theory of Clumps

It is assumed, in the work on the theory of clumps, that we have a population of objects or items among which at least some classes or groupings do objectively exist, but that we do not have any bases for precisely determining class membership requirements. There may, therefore, be many possible ways of grouping and many possible definitions of clumps. On the other hand, such diverse definitions must conform to the extent of some similarities of membership in the clumps that they define if in fact they do define any of the existing classes. Assuming further that we are given information about properties ascribable to various members of the population, it is theorized that useful clumps can be discovered by investigating similarity connections between pairs of items, such as the number of co-occurrences of specific properties. Thereafter, only these similarity connections are considered, and the connection matrix is used as the basis for trial partitions of the population into various possible subsets.

1/ Stiles, 1962 [575], pp. 10-12.

2/ Doyle, 1963 [162], p. 15.

3/ Hammond, et al. 1962 [270], p. 17.

In early work on clump definition, Kuhns of Ramo-Wooldridge 1/ proposed the use of a threshold value such that if a subset is a clump every pair of members in it has a connection strength equal to or greater than the threshold value and no member of the subset's complement has connections of more than threshold value to the members of the subset. In the more extensive investigations carried out by Parker-Rhodes and Needham (1960 [465], 1961 [434, 435, 464]), other clump definitions have been explored and specifically that of the "GR-Clump". This is defined as a subset of the universe such that all its members have a positive (or zero) bias to the subset and all non-members have a negative bias to it, where bias is defined as the excess (positive or negative) of the total connections of a member of the population to the members of the subset over its total connections to the members of the subset's complement, following the convention that the connection of the element to itself is taken as zero.

An iterative procedure for discovering GR-clumps can now be followed. This is based on an arbitrary initial partition of the given universe of elements into a subset and its complement. Then, since each element has a bias toward both the subset and its complement, differing only in sign, the biases of each element are computed. If the bias of a particular element is positive with respect to the subset, it is transferred to the subset if it is not already a member of it, and conversely if its bias is negative, it is transferred to the subset's complement if it is not already there. Each time a transfer is made, the biases are recomputed and the process is repeated until for a complete scan of all elements no further transfers can be made. The result is a GR-clump even though it may have no members or may contain all the elements of the universe. In such case, a further partition is made and the procedures are re-applied.

These GR-clump finding procedures have been applied to such diverse collections of items to be classified as archaeological artefacts and patients' symptoms as related to specific disease diagnosis. In the latter case, groupings were obtained that corresponded satisfactorily to certain specific disease syndromes, but no group was found corresponding to Hodgkin's disease where a great variety of symptoms typically occur. Needham comments: "I can scarcely conceive of a clump definition that would be likely to group these patients; I am unsure whether this is a reflection on clump theory or on Hodgkin's disease." 2/

In applications more directly related to documentation, some investigations have been made of the use of co-occurrence coefficients of index terms assigned to documents in order to form a connection matrix from which clumps were then derived (Needham, 1963 [431]). These experiments covered 342 terms occurring more than once in the index-term sets assigned to several hundred documents in the general subject field of machine translation. Computation of the matrix required 20 minutes of computer time and the 40 clumps found took 6-8 minutes each to find. Needham reports on the results as follows:

1/ See Kuhns, 1959 [336], and Needham, 1961 [435], pp. 20-21.

2/ Needham, 1963 [433], p. 46.

> "Evaluation of the results was unexpectedly difficult. The acid test is presumably the efficiency of the retrieval system embodying the grouping given by the program; but the efficiency of retrieval systems cannot be easily measured. An apparently simpler test would be to see if the clumps were intuitively satisfactory, i.e., were groupings that a classifier in his right mind could have made. This also was unsatisfactory because the groups are mostly rather large, larger in fact than classifiers ordinarily make, and were thus very difficult to judge. The test eventually adopted was to group the terms not distinguished by the clump classification, and look at these. Accordingly, for each term, a list of the clumps to which it belongs was prepared, and groups of terms were found which had all their clumps in common. These groups were quite small (2-6 terms) and could be studied easily. It turned out that some groups were ones of which a human classifier could have thought (e.g., words concerning suffix removal for machine translation came together) while others were quite justified by the documents concerned, but would never have been thought of a priori. For example, the group: "phrase marker, phoneme, Markov process, terminal language" was entirely justified by the... contents of the library. It is groups of the latter kind that represent a success for clump theory, for they function usefully in retrieval but in no way form part of the structure of thought... which the human classifier's work is likely to reflect." [1]

Still another application of the theory of clumps may be of use in the construction of thesauri (Sparck-Jones, 1962 [564]. Here the assumption is that rows of a correlation matrix can be formed for words giving other words which are synonymous with respect to meaning. The overlaps of the same word's occurrence in two or more rows can then be used to find clumps which are presumed to represent conceptual groupings.

Applications of clump theory to problems of mechanized documentation are also being investigated by Dale and Dale of the Linguistics Research Center, the University of Texas. [2] They have begun experimentation to derive clumps for the 90 clue words used by Borko and the 260 source-item computer abstracts used by both Maron and Borko. Preliminary results reported so far are principally limited to considerations of the associative networks between terms as derived from the structure of the clumps discovered by several clump definitions. Mention should also be made of the work of Meetham and Vaswani at the National Physical Laboratory, Teddington, England, looking toward the use of similar techniques for machine-generated index vocabularies, with preliminary emphasis on testing them against a "library" consisting of the propositions of Euclid's geometry. [3]

---

[1] Needham, 1963 [431], p. 285-286.

[2] Dale and Dale, an unpublished report dated February 1964, [147].

[3] National Science Foundation's CR&D report No. 11, [430], p. 137; and Meetham, 1963 [413].

5.3 Latent Class Analysis

Like the earlier work of Tanimoto, the latent class analysis approach of Baker (1962 [27]) to problems of automatic information classification and retrieval is at least to date theoretical rather than experimental in nature, and so will be considered only briefly here. Baker claims that the latent class model developed in the field of the sociological sciences for the determination of latent classes among individuals responding "yes" or "no" to items in a questionnaire would have attractive features for application to information categorization and search, because the model is based upon response patterns that are analogous to the presence or absence of clue words or phrases in documents and because the analysis yields an ordering ratio that could serve a function similar to the relevance weightings suggested by Maron and Kuhns.

This ordering ratio is the probability that a given pattern of clue words will occur in a document properly belonging to a particular latent class. The probabilities of the same pattern being generated by a document properly belonging to other classes are also provided, giving an uncertainty which Baker thinks justifiable because a "document could generate a given pattern of key words, yet not belong to the same area of interest as the majority of documents possessing the same pattern of keywords".[1] It should be noted, however, that the question of how to select appropriate clue words is begged[2] and that no computer programs are as yet available for carrying out latent class analyses.[3]

5.4 Examples of Other Proposed Classificatory Techniques

There are certain other document classificatory techniques that have been proposed and to some extent investigated experimentally. Trials of document clusterings based on co-citingness, co-citedness, or bibliographic coupling as compared with subject content groupings have, as noted above, been conducted both by Kessler at the M. I. T. Libraries and by Salton's group at Harvard.[4] Consideration of Doyle's work on word co-occurrence statistics has been deliberately deferred to a later section which covers his general "association map" approach. Similarly, several other investigations will be discussed in terms of potentially related research such as linguistic data processing.

Two particular examples of other suggested classificatory techniques for document grouping or classification are somewhat unusual, however. These are the methods proposed by Te Nuyl and by Lefkovitz (1963 [353]). Cleverdon and Mills comment on Te Nuyl's method as follows:

---

[1] Baker, 1962 [27], p. 518.

[2] Ibid, p. 517. Note also that the footnote states: "A referee of this paper has properly cautioned that the effectiveness of an information retrieval system may be due more to the appropriateness of the key words than the subsequent processing." See also Hillman, 1963 [272], p. 323: "Baker's theory, however, is based on interrelationships of key words, and thus constitutes an approach which is regarded with some suspicion by Farradane, who thinks that the real problem concerns the interrelationships of the concepts which key words denote."

[3] Baker, 1962 [27], p. 516.

[4] See Kessler, 1963 [320]; Lesk, 1963 [356, 357], and p. 30 of this report.

> "Te Nuyl...uses, as quasi-descriptors, word-sets chosen from the Oxford English Dictionary (e.g., any word falling between A-Ah) and relies on the subsequent correlation of terms to make sense of his seemingly bizarre choice." 1/

Lefkovitz is concerned with the so-called "automatic stratification" of a file in which both generic or associative relationships and exclusive partitioning is used to facilitate search. He claims:

> "... The exclusive partitioning implies a separation of descriptors into groups such that no two descriptors in a group co-occur in any given document description of the file. This arrangement presents the dissociative properties of the file, or forbidden combinations. When coupled with a superimposed display of the 'inclusive' or associative properties of the file a unique classification of the descriptors of this file results, which is based solely upon the association of the descriptors themselves within the document descriptions and not upon an arbitrary set of classes constructed by professional indexers." 2/

The purpose is to assist the searcher by warning him that if he chooses more than one descriptor from any one group as terms in his search request, there will be a null response from this particular file. However, the particular application considered involves a limited number of highly quantifiable or scalable "attribute-value" pairs, (for so the descriptors involved are defined), such as "Age-23", and "Hair-red". It is by no means obvious that comparable exclusive partitionings could be achieved for literature items or that the recomputations necessary as new items enter the file can be achieved on a practical basis.

6. OTHER POTENTIALLY RELATED RESEARCH

In this section we shall consider certain areas of potentially related research that may prove applicable to the improvement of automatic indexing techniques. First is the area of thesaurus construction and use, which in turn is somewhat related to the development of statistical association techniques, especially for "indexing-at-time-of-search" and search renegotiations. Natural language text searching will also be briefly considered, together with related research in the general area of linguistic data processing.

6.1 Thesaurus Construction, Use, and Up-Dating

The first area of potentially related research which promises improvements in automatic indexing procedures is that of thesaurus lookups by machine. There are several different possible definitions of the word "thesaurus" in the context of information storage, selection and retrieval systems. The first is that it is a prescriptive indexing aid, or authority list, serving the function of normalizing the indexing language, primarily by the use of a single word form for words occurring in various inflections, by the reduction of synonyms, and by the introduction of appropriate syndetic devices. The second definition relates to the intended function for the provocation and suggestion to the indexer or the searcher of additional terms and clues, and it follows the idea of word groupings related to concepts as in a traditional thesaurus like Roget's. The third

---

1/ Cleverdon and Mills, 1963 [131], p. 8.

2/ Lefkovitz, 1963 [353], Preface, pp. VIII-IX.

possible definition involves the special case of devices or techniques which display or use prior associations and co-occurrences or words, indexing terms, and related documents to provide a guide or suggestive indexing and search-prescription-formulation or renegotiation aid.

The idea of a mechanized authority list, following the restrictive first definition, has been proposed by a number of investigators [1] and has actually been used in computer programs as discussed for example by Schultz and Shepherd (1960 [532]), Shepherd (1963 [545]) and Artandi (1963 [20]). It is the second definition of thesaurus with which we shall be principally concerned. It is, as we have said, close to the conventional idea of such a thesaurus as Roget's. It is based on the hypothesis that patterns of co-occurrences of words in a new item or in a search request can be compared with patterns of prior co-occurrences, as given by a thesaurus "head", in order to expand, clarify, or pin-point "meaning" and thus provide a more effective indication of the true subject content. The third definition will be considered as falling within the more general scope of statistical association techniques, although as Giuliano points out, "a retrieval system embodying an automatic thesaurus thus qualifies as being 'associative'." [2]

The application of a thesaurus-like approach to indexing and searching problems is again an area in which Luhn is one of the earliest proponents. In January 1953, he proposed a new method of recording and searching information in which a special dictionary would be compiled for use in broadening the terms of a search request and in normalizing word usage as between various indexers (recorders) and searchers. Although he did not then use the term "Thesaurus" as such, he said in part:

> "The process of broadening the concept involves the compilation of a dictionary wherein key terms of desired broadness may be found to replace unduly specific terms, the latter being treated as synonyms of a higher order than ordinarily

---

[1] See, for example, "Summary of discussions, Area 5," ICSI, 1959 [578], p. 1263: "Two further complications arise from a mechanical index. Some articles might deserve as an indexing term a word not contained in the article. By an authority list, the product of the mechanized indexing procedure might have such additional words added to it. Again, an article might use a particular word but the vocabulary of the system might prefer another one. This also can be handled by a mechanized authority list.".

[2] Giuliano and Jones, 1962 [229], p. 4.

> considered. Translating criteria into these key terms is a process of normalization which will eliminate many disagreements in the choice of specific terms amongst recorders, amongst inquirers, and amongst the two groups, by merging the terms at issue into a single key term. However, the dictionary does not classify or index but maintains the idea of being fields... A specific term may appear under the heading of several key terms and if according to its application an overlapping of concepts exists then the term is represented by the several key terms involved..." [1]

In subsequent papers, Luhn has developed related ideas of a "family of notions" and "dictionaries of notional families". [2] In particular, he emphasizes that for automatic indexing, by contrast with automatic abstracting, consideration should be given to the normalization of variations in author-chosen terminology: "It will be necessary for a machine to resolve variation of word usage with the aid of a device the functions of which resemble a dictionary at one level and of a thesaurus at another level of requirements." [3]

The first issue of the National Science Foundation's compendium of project statements, "Current Research and Development in Scientific Documentation", which appeared in July 1957 [430] reported several projects of interest in terms of thesaurus construction and use, [4] namely: (1) work by Luhn at IBM involving the establishment of a thesaurus to facilitate encoding of items whose texts would be available in machine-usable form, (2) work by Bernier and Heumann at Chemical Abstracts Service looking toward the development of a technical thesaurus, (1957 [57]), and (3) an approach to mechanized translation proposing to use a mechanized thesaurus at the Cambridge Language Research Unit. This latter project incorporated the ideas of Masterman and her associates from about 1956 on (Halliday 1956 [249], Masterman, 1956 [403]; Joyce and Needham, 1958 [305]), to apply the principle of checking co-occurrences of text words against thesaurus "heads" to which they belonged, in order to resolve homographic ambiguities and thus achieve more idiomatic translation by machine.

For the ICSI Conference in 1958, Masterman, Needham and Sparck-Jones prepared a paper discussing analogies between machine translation and information retrieval, and recapitulated the arguments of Needham and Joyce for the use of a thesaurus in the formulation of search requests, as follows:

> "If a large number of terms are used to describe a document, the existence of synonyms is likely: in a system such as uniterm no attempt is made to bracket the synonyms, which means that a request will produce only the document described

---

[1] Luhn, 1953 [383], p. 15.

[2] Luhn, 1959 [371], p. 51, 1959 [384]; 1957 [385], p. 316.

[3] Luhn, 1959 [384], p. 12.

[4] National Science Foundation's CR&D Report No. 1, [430], pp. 21, 6, 4.

> in identical terms and not in synonymous ones. If the existence of synonyms is avoided, by using a small number of exclusive descriptors, the description of a document in terms useful for retrieval is more difficult, also it is equally difficult to relate a request to the description of documents. A further difficulty is that descriptions only list the main terms, and take no account of their relations to one another. The C. L. R. U. experiments being carried out make use of a thesaurus, a procedure through which it is hoped that these difficulties will be avoided and that a request for a document although not using the same terms as those in the document will produce that document and others dealing with the same problem, but described in different, though synonymous, terms." [1]

In general, the use of a thesaurus to constrain variations in word or term usage (as in our first definition, a mechanized authority list), to reduce synonymity, to resolve homographic ambiguity, to provoke and suggest additional terms or ideas to indexer and to searcher alike, is related to the improvement of automatic indexing procedures in precisely the same sense that its use would be effective in any indexing system whatsoever. In another sense, however, the construction and use of the thesaurus is related to linguistic data processing by machine in another way. Garvin suggests:

> "...One may reasonably expect to arrive at a semantic classification of the content-bearing elements of a language which is inductively inferred from the study of text, rather than superimposed from some viewpoint external to the structure of the language. Such a classification can be expected to yield more reliable answers to the problems of synonymy and content representation than the existing thesauri and synonym lists, which are based mainly on intuitively perceived similarities without adequate empirical controls." [2]

This is with respect to the recognition that the machine itself can be used to compile and construct the thesaurus. While Luhn in some of his 1957-8 proposals still considered the compilation and organization of a thesaurus to be primarily a matter of human effort, he nevertheless pointed out that: "The statistical material that may be required in the manual compilation of dictionaries and thesauri may be derived from the original texts in any desired form and degree of detail." [3] De Grolier makes the complementary statement that the Luhn techniques should "considerably facilitate" the preparation of thesauri. [4]

Even more importantly, the computer can be used for periodic up-datings and revisions. The work on the FASEB index-term normalization procedures involved early recognition of the need to "educate the thesaurus" by examining print-outs when no matches occurred and providing a continuous process of amendment. [5] Computer-maintained statistics of word and term usages are closely related to possibilities for construction and revision of a mechanized thesaurus, as again Luhn has suggested.[1] Schultz suggests that machine records should be maintained of what thesaurus terms are actually used for indexing and searching, the frequencies of term usage, the co-occurrences, the number of items described by particular combinations of terms and the like.[2]

---

[1] Masterman, Needham, and Sparck-Jones, 1958 [405], p. 934-935; Needham and Joyce 1958 [305].

[2] Garvin, 1961 [224], p. 138.

[3] Luhn, 1959 [354], p. 12.

[4] De Grolier, 1962 [152], p. 132.

[5] Shephard, 1963 [545], p. 392.

The potential combinations of natural text processing, automatic indexing, and thesaurus construction and updating are stressed in many current programs. For example, Eldridge and Dennis discuss:

> "Indexing by machine from natural text in a fully automatic system, in which statistical analysis of the words is employed as a device for (a) building automatically a 'concept' thesaurus, (b) indexing incoming documents with reference to the thesaurus, and (c) continuously revising the thesaurus to reflect new word usages in currently incoming documents."

Similarly, Giuliano and Jones suggest that given a term-term statistical association matrix, a transformation can be arrived at with a unit vector assigning value only to index term Z that ranks every other index term according to degree of association with Z, then by listing the higher ranked terms for each term Z, "a 'thesaurus' listing can be obtained completely automatically." [4]

6.2 Statistical Association Techniques

A special definition of the word "thesaurus" might, as we have noted, include the development of devices and techniques which either automatically or by man-machine interaction serve to suggest the amplification of a set of index terms. We shall briefly consider here both devices that visually display associations between words, terms, and documents [5] and techniques for machine use of coefficients of correlation for prior co-occurrences in a collection of word-word, word-term, term-term, term-document, and document-document associations, the statistical association factor technique as first developed by Stiles.

---

[1] Luhn, 1957 [385], p. 316: "Provision should be made to register the number of times each word is looked up in the index and the number of times each family number has been used for encoding. Such a record would be an indispensable part of the system for making periodic adjustments based on the usage of words or notions as mechanically established."

[2] Schultz, 1962 [529], p. 104.

[3] Eldridge and Dennis, 1962 [183], p. 6.

[4] Giuliano and Jones, 1962 [229], p. 12.

[5] It should be noted that Tabledex, the Scan-Column Index, and similar tools provide to some extent a display of prior associations between index terms. (See pp. 25-27 of this report.) Thus Cheydleur (1963 [115], p. 58) remarks: "Ledley.. has focussed on inter-item concepts in designing his economical TABLEDEX arrangement for displaying the connectivity of index terms and related file items."

6.2.1 Devices to Display Associations: EDIAC

The interest aroused among some documentalists by the provocative idea of a "Memex" to record and display associations between ideas as proposed by Bush in 1945 ([93]) led to specific attempts at Documentation, Inc. in the 1950's to develop a device which would incorporate at least the associations between indexing terms assigned to documents and between documents with respect to their sharing of common indexing terms (1954 [157], 1956 [155, 156]). The first approach to this objective, as reported by Taube, was the idea of a manual dictionary of terms arranged in alphabetical order, with a "page" reserved for each and every indexing term used for any document in the collection. On each page would be listed all other terms that had co-occurred with that term in the indexing of one or more documents. Another idea was to display associations of terms used in a collection through the "superimposition of dedicated positions in a set of cards or plates..." 1/

Subsequently, an actual device to demonstrate a system for display of term-term, term-document, and document-document associations, was built under an Office of Naval Research contract. 2/ The demonstration model contained a vocabulary of 250 terms which had been used in various combinations to index 100 reports. Interconnections in an electrical network provided the associational linkages. A display panel was provided with symbol-indicators which could be lighted up to identify particular terms and particular report numbers.

This EDIAC device (for Electronic Display of Indexing Association and Content) was intended for use both in guiding an indexer to either the extension or refinement of his initial choice of indexing terms and in assisting the searcher. It was claimed that the operation of such a device would be extremely simple. Thus:

> "For the index question the searcher selects any term in which he is interested and applies a voltage. He is told instantly the number of the reports dealing with that subject. Putting voltage in at any term also lights all other terms associated with the first term..." 3/

A later analog device, ACORN, will be discussed below in connection with the work of Giuliano and associates, at Arthur D. Little, Inc.

6.2.2 Statistical Association Factors - Stiles

The name of H. Edmund Stiles, like those of Luhn, Baxendale, Maron, Swanson, Edmundson and Wyllys, is generally associated with pioneering innovations in those areas of mechanized documentation which are directly related to the use of high-speed computer capabilities. While Stiles' work has been directed primarily to problems of search prescription formulation and renegotiation based on the results of preliminary search, he has specifically recognized that the use of statistical word association techniques in searching operations can provide a logical corollary to automatic indexing procedures. Thus:

1/ Taube et al, 1954 [599], p. 102.

2/ It is described and illustrated in Taube et al, 1956 [599], p. 63 ff.

3/ Documentation, Inc. 1956 [156], p. 7.

> "Automatic indexing, based on the relative frequency of words used in a document, produces a partial vocabulary of the content words used to express its subject. Retrieval can then be accomplished by expanding the request vocabulary... This method tends to overcome the deficiencies and inconsistencies inherent in the use of terms derived automatically from a text." 1/

Conversely, Stiles also points out the possibility that the results of automatic derivative indexing procedures, extracting indexing words from the documents directly, might prove a more realistic or reliable basis for the development of his word co-occurrence correlation data than do the Uniterms assigned by human indexers. 2/ The work of Stiles has also stressed the importance of two factors that may well be critical for the improvement of automatic indexing techniques. These are, namely, the consensus of prior human indexing and the consensus of subject coverage of a particular collection. 3/

In his experimental investigations, Stiles began with an existing collection of approximately 100,000 items which had previously been indexed, over a period of time, with a Uniterm indexing vocabulary consisting of about 15,000 terms. The objective of the experiments was to determine how, given a specific search request, a more effective "net to catch documents" 4/ could be generated and how the responding items might be ranked in order of their probable relevance to the request.

The statistics of co-occurrence of terms used to index the same documents were first obtained. A modified chi-square formula was then applied to determine relative frequencies of use of co-occurring terms. 5/ Patterns of term co-occurrence could then be derived in the sense of term-profiles which show, for each term, the more significant of its associational values of pairing with other terms in the collection. The actual procedure for using these term-profiles in search prescription formulation and in document selection involves several steps, generally as follows: 6/

1/ Stiles, 1962 [573], pp. 12-13.

2/ Stiles, 1961 [572], p. 205.

3/ Stiles, 1962 [573], p. 6 and 1961 [572], pp. 273, 277.

4/ Stiles, 1961 [572], p. 192.

5/ In general, we shall not be concerned with the precise mathematical formulations. It is to be noted that in a recent report Giuliano and his colleagues have reviewed a number of the various mathematical formulas proposed in the literature for the computation of word, term, and document associations, including those of Parker-Rhodes and Needham, Maron and Kuhns, Stiles, Salton, Osgood, Bennett and Spiegel (Giuliano et al, 1963 [230], Appendix I).

6/ Stiles, 1961 [571], pp. 273-275.

1. For each term in the initial formulation of a search request, the appropriate term-profile is obtained, which gives weighted values for those other terms that had significantly co-occurred with it.

2. The profiles of each term in a multi-term request are compared and those additional terms common to all or a specified number of the profiles are selected and added to the initial set.[1]

3. The "first generation" terms resulting from step 2 are next treated as though they also were request terms, and steps 1 and 2 are repeated for them.

4. A selection is made from some reasonable proportion of the profiles associated with the first generation terms to produce the "second generation" terms.[2]

5. The expanded list of search terms is then compared with the index terms assigned to each document in the collection, and whenever a match is found the weight of the request term is assigned to the matching document term. These weights are then summed to provide a numeric measure of probable document relevance to the original request.

6. Documents responding to the expanded request are printed out in the order of document relevance scores.

Some experiments have been made using a computer program which accepts up to 300 weighted terms in an expanded request vocabulary. Representative results have been reported, in part, as follows:

> "...We asked a qualified engineer to examine these documents and specify which were related to 'Thin Films' and which were not... This engineer was not familiar with our project...yet...we found a remarkably high correlation between his evaluation and the document relevance numbers... We then checked to see how the documents containing information on 'Thin Film' had been indexed. We found that the first five documents on our list had been indexed by both 'Thin' and 'Film'. Three more documents had been indexed by 'Film' alone, and other related terms. Two documents had not been indexed by either 'Thin' or 'Film', but only by a group of related terms, yet they contained information on 'Thin Films' and had a high document relevance number. By using association factors and a series of statistical steps, easily programmed for a computer, we were thus able to locate

---

[1] These are called "first generation terms" and tend to reflect only statistical associations without including synonyms and near-synonyms which, over the course of time, have occurred in the indexing vocabulary.

[2] Stiles, 1961 [571], p. 274: "Among these we find words closely related in meaning to the request terms." An example given in Ref. [572], pp. 200-201, is the derivation of 'weathering,' 'fungicidal', 'deterioration', and 'preservatives' as second generation terms when the initial request included the terms 'plastics', 'fungus', 'coating', and 'tests'.

documents relevant to a request even though the documents had not been indexed by the terms used in the request." [1]

In another case, which was analyzed in detail, a request profile of 26 terms that had been intuitively weighted by the customer resulted in the machine listing of 246 presumably responsive documents. Of these, 81 documents were of primary interest to the customer, and an additional 78 were of secondary interest to him. [2]

The statistical association technique as proposed by Stiles has also been investigated at the Datatrol Corporation, with particular reference to the field of legal literature (Hammond et al, 1962 [251]). About 350 documents in the field of Federal public law were indexed in cooperation with George Washington University, using a vocabulary of 680 index terms. A computer program was written for the IBM 7090 that can accommodate a 1200 x 1200 matrix to calculate the Stiles' association factors. Trials were made of various thresholds to determine which other terms were sufficiently high in association strength to a particular term to be selected for that term's profile.

Given the generation of the term profiles, a less sophisticated computer such as the 1401 can be used for the expansion of request terms and the actual conduct of searches. Such a program was demonstrated at the Annual Meeting of the American Bar Association, August 1962, with running of "live" requests suggested by jurists and with what are claimed to be "highly gratifying results". A point of interest relates to the question of updating of term-profiles and other statistical association factor data. Hammond, et al report:

> "The term profiles were generated a total of three times in the course of the pilot study, making it possible, to some extent, to assess the effect of vocabulary growth. Judging from this limited experience, it appears that a bi-monthly, or perhaps even quarterly, recompilation of term profiles should be sufficient for a mature collection." [3]

6.2.3 The Association Map - Doyle and Related Work at SDC

The name of Doyle is again that of an early and prolific investigator and innovator in the field of mechanized documentation and linguistic data processing. One of his provocative suggestions is generally known, in his own terminology, as that of "semantic road maps for literature searchers" or an "association map" technique. As a matter of convenience, we have chosen to consider this suggestion and a variety of related work under the general heading of the association map technique, [1] although passing reference has been made to some of Doyle's suggestions and findings elsewhere in this report.

---

[1] Stiles, 1961 [577], pp. 198-199.

[2] Stiles, 1962 [573], p. 9.

[3] Hammond et al, 1962 [251], p. 6.

Beginning in 1958 (Doyle, 1959 [168]) information retrieval projects at the System Development Corporation have had, among other objectives, that of developing ways to use computers in the processing and interpretation of natural language text. By February of 1959, a computer program was already in operation that could search fragments of about 100 words of keypunched text, match input words against a pre-established clue word selection list (i.e., an inclusion dictionary) and substitute a short encoded form to be used for subsequent search. Processing of keypunched abstracts using this program involved computer time at the rate of four abstracts per second.

Other features of this text compiler, and of subsequent text processing programs developed at SDC, enable the making of frequency counts and other statistical measures. Such features are then used for the investigation of, for example, word-word, word-document, and word-subject associations, looking toward the determination of answers to such questions as: "Do subject words have distribution characteristics within a library that a computer program can detect?" [2]

Doyle's investigations of word co-occurrences have included hypotheses and tests of various probabilistic measures in terms of observed frequencies, in terms of "boing!" words (so-called because of the mental sound effect they elicit), [3] in terms of adjacent word pairs and affinities between particular nouns and particular adjectives, [4] and in terms of distinctions between frequency (the total number of times a word appears in a given library corpus) and prevalence (the total number of items in which a particular word appears). [5] He has also stressed distinctions between adjacent words and high correlations for words that are not closely positioned together in text, as follows:

---

[1] Compare Doyle himself, 1962, [163], p. 383: "Swanson and others have offered thesauri of synonyms and related terms...(to assist in indexing or search processes)...An association map is, in a sense, an extension of this solution; it is a gigantic, automatically derived thesaurus. Confronted by such a map, the searcher has a much better 'association network' than the one existing in his mind, because it corresponds to words actually found in the library, and, therefore, words which are best suited to retrieve information from that library." See also Wyllys, 1962 [651], p. 16: "L. B. Doyle (1961) has invented a fascinating search tool which seems to us to belong at a level intermediate between automatic indexes and automatic abstracts; i.e., a possible search method might be to have the computer scan automatic indexes and compare the index terms therein with the request, then obtain the possibly pertinent documents and display their association map for the user to examine..."

[2] Doyle, 1959 [168], p. 6.

[3] Doyle, 1959 [165], p. 5.

[4] Doyle, 1961 [169], p. 12; 1959 [165], p. 16.

[5] Doyle, 1962 [163], p. 380.

"We have also perceived that two different cognitive processes seem to be responsible for each type of correlation, one (adjacent correlation) involving the habitual use of word groups as semantic units, and the other (proximal correlation) having to do with the pattern of reference to various aspects of that which is being discussed. We can call the statistical effects, respectively, 'language redundancy', and 'reality redundancy'. Such a resolution of statistical effects is full of significance for information retrieval because it appears likely that reality redundancy can vary greatly from one science to another, whereas language redundancy, a universal property of talking and writing, is relatively invariant." [1]

With respect to the "semantic roadmap" or "association map" technique itself, Doyle's suggestion is that various measures of word and index term cross-associations may be applied to the generation of graphic displays of both types of co-occurrence relationships. Because of the variety of, in particular, the "proximal" correlations, it is assumed that the literature searcher should be given a display in which the representation of the assemblage of the varied relationships is two-dimensional rather than one. [2] An example is given, based upon computer processing of 600 abstracts of SDC internal reports to find intersections between 500 topical words, of associational connections for the word "output". This was generated by selecting the eight words most strongly correlated in the data with "output", such as "manual" and "radar", and then finding three other words highly correlated with each of these and also correlated with "output" itself. From the initial graph, it is further shown that item surrogates might be generated by word selection rules applied to documents to pick up, for example, "New York Air Defense → system → data → outputs → D.C. [3]

Continuing related work by Doyle and others at SDC has included various experimental studies of "pseudo-documents" consisting of lists of the twelve most frequently occurring words in 100-item samples of abstracts in various subject fields (Doyle, 1961 [161]). Of special interest in terms of potential improvements and modifications to machine indexing techniques are studies, based on similar lists, looking to the separation of words that may have been used in several different senses, i.e., the detection of homographs by statistical means (Doyle, 1963 [171]). More recent investigations by Doyle involve considerations of differences between word-grouping and document-grouping techniques and of possibilities for use of hybrid methods.

6.2.4 Work of Giuliano and Associates, the ACORN Devices

A program directed toward the design of "an English command and control language system" under an Air Force contract with Arthur D. Little, Inc., involves several interrelated aspects of natural language text processing, use of statistical association factors in search, man-machine interaction during search, and display of associational relationships by means of analog network devices. In this program and in related research, Giuliano and his associates are convinced that:

---

[1] Doyle, 1961 [169], p. 15.

[2] Doyle, 1962 [163], p. 379.

[3] Doyle, 1961 [169], pp. 24-25.

"Automatic index term association techniques are needed to improve the recall of relevant information, to enable indexers and requestors to use language in a more natural manner, and to enable retrieval of relevant messages which are described by different index terms than those used in the inquiry." [1]

For the most part, the work to date has been directed to "associative retrieval" of messages limited to single sentences of English text, and to the search phases of a proposed system.

In the case of a corpus consisting of 230 sentences from a single text, a partially automatic indexing method was used. The text was first processed against a modified version of the Harvard Multipath Syntactic Analysis computer program and the resulting analyses were manually screened to select a unique, correct analysis for each sentence. Next, approximately 500 words, those that had been marked "noun" by the syntactic analyzer, were listed out and these in turn were manually screened to provide an "inclusion list" of 273 words. Sentences were then "indexed" with respect to which of these selected words they contained. Word associations were computed both in terms of co-occurrence within a sentence and of co-occurrence in syntactic structures.

Retrieval tests were then applied using both computer programs and the analog device, and evaluations were made on the basis of examining sentences selected in order of machine-ranked relevance and of comparisons of word lists associated with a given search term against association lists for another term picked at random. It is noted that, "although quantitative conclusions cannot be drawn", the results support the conclusion that: "Items retrieved due to automatically-generated associations tend to be more relevant than is explainable on a chance basis." [2]

The "request reformulation" retrieval program has also been used to generate term profiles from a collection of approximately 10,000 documents (previously indexed with at least 6 terms from a selective term vocabulary of 1,000 terms) which have then been compared against lists provided in the entries for corresponding terms in the Thesaurus of ASTIA Descriptors, Second Edition. The machine-produced association lists, at least for those words occurring relatively frequently in the corpus, appear to give thesaurus entries that are extensive, specific, and intuitively acceptable, and of high quality, especially with respect to listings of synonyms as well as factually related words. [3]

The development of the ACORN (Associative Content Retrieval Network) devices has provided additional tools for testing and display (1962 [229], 1963 [227, 304]). These devices are networks of passive resistance elements. Each word or index term and each sentence (240 by 230 in ACORN-IV) are represented by terminals interconnected by resistors with conductance equal to the connection strength, and with "leak" resistors providing for various normalizations that may be applied to compensate for word or sentence frequency factors. These devices differ from the earlier EDIAC in the variable weightings provided, in the normalizations that may be applied, and in multipath interconnections.

---

[1] Giuliano 1962 [228], p. 10.

[2] Giuliano et al, 1963 [230], p. 47.

[3] Ibid, pp. 57-58.

When, for example, currents are applied at some of the word terminals, the voltages appearing on any of the other word terminals depend on the strengths of association between these words and the input words via all direct and indirect paths. The responses of sentence terminals to the input words of a query similarly depend upon how strongly a sentence is connected to these words and how strongly it is connected to other words which in turn are strongly connected to the query words. It is to be noted further that:

> "Pulling out or cutting a few randomly selected wires in an ACORN generally has a surprisingly small effect... This insensitivity is of course, explainable in terms of the multiplicity of indirect and redundant association paths which remain intact when a direct path is severed... It... suggests that the retrieval process can indeed be made insensitive to minor variations in indexing." [1]

In addition, there are intriguing possibilities for imposing a "viewpoint" with respect to a search by injecting bias currents. Thus if only non-"Air Force" jet plane items were desired, the "Air Force" items could in effect be grounded out. If there were no jet items in the collection other than those which were also Air Force items, these would be indicated as responsive, but largely they would appear only if this should be the case. Some words used have some connection to almost all other words, but these have little effect in the system and the hardware thus tends to compensate for the high frequencies of very general words.

6.2.5 Spiegel and Others at Mitre Corporation

Bennett and Spiegel, reporting at the Symposium on Optimum Routing in Large Networks, IFIP Congress-1962, [2] consider modifications to formulas for the calculation of statistical association factors which will normalize against such influences as frequency of word occurrences, relative word position within a string of words, and string length. This work has been carried forward at the Mitre Corporation in a program for developing procedures to encode various statistical properties of messages or documents and to use these codes for message routing and retrieval.

Differences between this approach and those of Maron and Kuhns, Stiles, and Doyle, relate primarily to the questions of how best to normalize. The objective is closely similar: to use associational weighting so as to provide, in response to a query, output of documents or messages ranked in order of probable relevance to the query.

---

[1] Giuliano and Jones, 1962 [229], p. 22.

[2] See Juncosa, 1962 [306], especially paper 4, E. Bennett and J. Spiegel, "Document and message routing through communication content analysis", pp. 718-719.

Additional features include provision for the matrix of coefficients of association to change with time or with deliberate manipulation to improve performance. Thus:

> "Each normalized cell weight... rises and falls with time as each specific association increases or decreases in relative frequency. In this way, the matrix memory of associations changes with time, maintaining a cumulative pattern of associations reflecting one statistical characteristic of messages fed into it in the past...
>
> "In addition to this adaptive characteristic of changing memory with time and with changing inputs, the matrix is also readily subject to formal education. Any specific cell weight can be strengthened by repeatedly reading into the matrix memory the specific strings that contain the desired associations. For example, by introducing the strings is am, is are, am is, am are, and are am, we can increase the statistical tendency of the tokens is, am, and are to be associated." [1]

Experimental results have been obtained for a corpus of 500 bibliographic entries contained in DDC's Title Announcement Bulletin. In the case of a three-term query, 40 items were selected and ranked in probable relevance order, with selection based on a particular relevance score value threshold. The investigators then reviewed the abstracts of all 500 items and rated them as to relevance with respect to the query. Seven additional items were found, of which three would have been machine-selected with a less stringent selection threshold. For the remaining four, it is reported that they "were poorly indexed and could have been judged not relevant by a human who depended upon the descriptor string only, as the matrix did, rather than upon review of the abstracts." [2]

6.3 Clues to Index-Term Selection from Automatic Syntactic Analysis

Several of the organizations and research teams most active in the investigation of linguistic data processing techniques, especially for automatic indexing, extracting and search renegotiation applications, are actively considering the use of clues derived from automatic syntactic analysis to improve criteria for machine selection of "significant" words, phrases, and sentences from raw text. Such approaches, in general, however, are subject to the limitations of non-availability of sufficient corpora of text in machine-usable form, in the first place, and, even more importantly by the non-availability of satisfactory computer programs for complete syntactic analysis up to the present time. [1] In terms of the state-of-the-art of automatic indexing, therefore, we shall not consider these approaches as more than indications for future research. A few suggestive examples are discussed briefly below.

---

[1] Spiegel et al, 1963 [566], p. 17.

[2] Ibid, p. 34.

The multi-pronged attack on mechanized information selection and retrieval problems headed by Salton and his associates includes the exploration of tree structures, to represent both the relationships between terms in a classification schedule or indexing term vocabulary and the representation of the results of automatic syntactic analyses of natural language text. It is proposed, then, that computer programs can achieve transformations of the syntactic trees representing word strings in the original text into simplified, condensed structures with normalized terms and can compare these trees with the classificatory trees (Salton, 1961 [516]). Manipulation of such trees together with appropriate dictionaries or thesauri can result, for a given proposed index term, in the finding of a preferred term for a particular system, or a set of synonymous terms, or sets of all terms in which the given term is included, and the like.

Anger considers some of the problems involved in complete syntactic analysis of texts with the objective of identifying the total network of relationships expressed and implied, as proposed by Lecerf, Ruvinschii, and Leroy, among others, of the Research Group on Automated Scientific Information (GRISA), EURATOM. Assuming that computer programs for syntactic analysis are or will be available, he suggests that simplifications may be obtained by determining only the basic relations that are indicated by direct syntactic dependencies or by linking words, (Anger, 1961 [15]).

A specific program for automatically extracting syntactic information from text has been studied by Lemmon (1962 [354]). The possibilities for combining dictionary lookups, word suffixes as indicators of syntactic role, and predictive syntactic analysis for text processing have also been further explored by Salton himself (1962 [518], 1963 [519]). A variety of word and document association techniques and of synonymous word and phrase groupings which serve to "clue" the selection of a subject heading are also being investigated by members of the Harvard group and guest investigators.

---

[1] Major difficulties have to do with limitations both upon grammars and vocabularies so far tested and with ambiguities and the number of alternative parsings generated. See, for example, Bobrow, 1963 [68]. Kuno and Oettinger, 1963 [341] and Robinson, 1964 [502]. Bobrow provides a survey of syntactic analysis programs as of 1963, noting limitations or restrictions on each. He reports, for example, that available programs to compute word classes are not always correct in the class assignments made and that analysis systems are not complete unless they provide means for distinguishing between "meaningless strings and grammatical sentences whose meaning can be understood". He concludes: "Until a method of syntactic analysis provides, for example a means of mechanizing translation of natural language, processing of a natural language input to answer questions, or a means of generating some truly coherent discourse, the relative merit of each grammar will remain moot." ([68], p. 385) Robinson ([502], p. 12) says of sentences which can be parsed correctly, that they are: "Usually short sentences with no complicated embeddings of relative clauses and few participial or prepositional phrase modifiers. These include the basic sentences that most grammars are equipped to handle and that adult writers seldom produce."

Another partial approach to applying syntactic analysis techniques to automatic indexing is based upon syntactic word-class recognitions. Giuliano and his associates at Arthur D. Little, Inc., (1963 [230]), have investigated on a small-scale basis the use of the Kuno-Oettinger programs developed at Harvard for this purpose (Kuno and Oettinger, 1963 [340]). The broad program of information and language data processing research at System Development Corporation specifically includes investigations of structural patterns of sentences at the syntactic level and also of semantic factors such as the studies of polysemy and homographic ambiguity by Doyle, Wasser, and others. Borko reports:

> "... We... are analyzing actual written text for multiple meanings... The data for this study were drawn from the corpus of 618 psychological abstracts. Tabulations of frequency of paired and single word listings were used. A number of corpus-derived word frames have been prepared. Although this research is still in its early phase, we feel that we have made a good start on the problems of semantic analysis." [1/]

In Czechoslovakia, at the Karlova Universita, both statistical and semantical methods for' automatic abstracting are reported as being under consideration.[2/]

Other examples of proposals for the use of syntactic analysis techniques for the improvement of automatic indexing products include those of Spangler, Levery, Plath, Thorne, and Climenson and his colleagues at RCA, as well as the suggestions of those whose interests in automatic syntactic analysis have been primarily directed to problems of machine translation or more general problems of linguistic analysis. Hays, for example, although principally concerned with MT, indicates that the methods for determining phrase structures have obvious applications to the automatic determination of categories useful in the indexing of documents. [3/]

An existing GE-225 computer program for KWIC-type indexing from both titles and abstracts at General Electric's Phoenix Laboratories is being extended to incorporate word analysis features taking into account both syntactic and semantic aspects of a given line or sentence of text. [4/] Levery provides an example of similar directions being explored in European research, more generally oriented toward linguistic considerations as such than to machine-derivable criteria (largely statistical to date), which seek to combine the benefits of both human and machine processes by way of automatic syntactic analyses. He claims, for example, that:

[1/] Borko, 1962 [75], p. 6.

[2/] National Science Foundation's CR&D report, No. 11 [430], p. 123.

[3/] See Hays, 1961, [258], p. 13: "... Two broad problems on which work is just beginning at RAND: grammatic transformations and distributional semantics. The latter problems are especially important for automatic indexing, abstracting, and text searching." See also de Grolier, 1962 [152], p. 137.

[4/] National Science Foundation's CR&D report No. 11 [430], p. 21.

> "...The study of the position of keywords in the text and the syntactical relationship which exists among them will show the way to automatic abstracting and the use of more sophisticated retrieval systems." [1]

Plath suggests that, given a computer program to perform the parsing and syntactic diagramming of a text sentence, the results can serve quite usefully to augment the selection criteria based initially on statistical techniques, such as word-frequency counting. He says, for example:

> "Another possible application of the outputs of the sentence diagramming program is their employment as an aid in language data processing for purposes of information retrieval, particularly in systems for automatic literature abstracting of the sort proposed by Luhn (1958). The feature of the tree diagrams which is pertinent here is that the main components of a clause, including subject, verb and object, always correspond to the 'main topics' in an outline, and are therefore located at the upper levels of the tree. When the words on these upper levels are considered apart from the lower-level structures which modify them, they often summarize the content of the sentence in a sort of 'newspaper headline' or 'telegraphic style'." [2]

The problems of multi-level selection, or screening, such that machine programs for selection of the most probably significant words, phrases, or sentences can be focussed upon the most probably content-relevatory areas of text, are treated here, as also by Salton, in the sense of a cutting-off at a given depth in the analyzed syntactic structure. [3] A potentially important contribution to the future prospects for automatic indexing, however, lies in the "discourse analysis" and "transformational linguistics" approach of Harris (1959 [254]), where condensations and concentrations of similarities and differences of topical interest may hopefully be achieved.

Harris himself suggested, at least as early as 1958, applications of his approach to both automatic indexing and abstracting. A goal of the analyses he has proposed is to identify 'kernels' of linguistic expression, having first, by various transformations such as from passive to active voice, brought together different ways of saying the same thing. He then suggests not only machine operations to normalize by application of his transformational rules but also to determine:

> "...Which kernels have the same centers in different relations (e. g., with different adjuncts), and other characterizing conditions. The results of this comparison would indicate whether a kernel is to be rejected or transformed into a section... of an adjoining kernel, or stored, and whether it is to be indexed, and perhaps whether it is to be included in the abstract." [4]

---

[1] Levery, 1963 [359], p. 236.

[2] Plath, 1962 [474], pp. 189-190.

[3] See also Thorne, 1962 [605], p. v: "The approach followed requires that the computer itself syntactically analyse input text in order to convert it into special form called FLEX, which preserves only that syntactic information which is useful for data retrieval purposes."

[4] Harris, 1958 [254], p. 949.

Certain difficulties are self-evident. Consider, for example, the admittedly hypothetical text which might refer in various places to the "dissolute, disreputable, illiterate, elder Lincoln" (underlining supplied) and which might be so processed by machine as to imply that Lincoln the son was, although also President of the United States, "dissolute," "disreputable," "illiterate," and "elder." These, however, are difficulties that plague almost any machine processing of natural language text.

Climenson, Hardwick, and Jacobson have explored some of the possibilities of the Harris approach in experimental computer programs for the RCA 501 (1961 [133]). Specific features of these programs include:

1. Establishment of the syntactic class or classes to which a given word can belong, by dictionary lookup.
2. Investigations of sentence structure and context in an attempt to resolve the homographic ambiguities involved when the same word may function either as a noun or a verb.
3. Isolation and marking of sentence segments, such as noun phrases, prepositional phrases, adverbial phrases, and verb phrases.
4. Identification and marking of segments -- clauses or degenerate clauses.

On a very preliminary basis, a limited set of word and phrase deletion rules were set up and several sample documents were processed against them, yielding reductions to about 35 percent of the original text. These results suggest that "syntactical filtering criteria" might be applied to the improvement of modified derivative indexing techniques, such as the word-frequency counting techniques, either by deleting syntactically insignificant parts of selected sentences, or by counting identical phrases rather than words. The investigators conclude, however, that:

> "A formal linguistic approach to the problems of natural language processing promises to yield results vital to the success of automatic indexing and data extraction. But the work required in such an approach will be quite arduous; a long-range man-machine effort will be required to formulate practical machine programs for indexing and abstracting." [1]

A final special case of linguistic data processing involving syntactic analysis is that of Langevin and Owens. They claim:

> "A critical review of the analysis work done on the Nuclear Test Ban Treaty by use of the Multiple Path Syntactic Analyzer demonstrates that such a device can, even at present, provide a powerful technique for the systematic discovery of ambiguities in treaties and other documents. Because the analyzer operates without bias from the overall context of the document, it may sometimes be possible for it to discover ambiguities that would easily escape a human reviewer who knows what the document is 'supposed to say'." [2]

---

[1] Climenson et al, 1961 [133], p. 182.

[2] Langevin and Owens, 1963 [346], p. 26.

6.4 Probabilistic Indexing and Natural Language Text Searching

As in the case of automatic indexing proposals based upon automatic sentence extraction techniques, machine searching of full natural language text has been suggested as a basis for, at least, automatic derivative indexing. We have remarked previously that the machine use of complete text can only be considered to be "indexing" in a very special sense, that it is subject either to the non-availability of suitable corpora already in machine-usable form or to high costs of conversion to this form, and that too little is yet known of linguistic analysis and searching-selection strategies effectively applicable to natural language materials. Various examples of corroborating opinion, other than those previously cited, are as follows:

> "Machine searching is superb if it is known exactly how to describe the object of search, and if one could know how to choose from among many possible searching strategies. I doubt if any one is yet in this comfortable position with respect to machine searching of text." 1/

> "The most effective programs in automatic linguistic analysis have served only to illustrate how really complex is the structure of the language, and how far removed the present state of the art is from any system which might be useful in practice."2/

> "The recognition of words involves only the matching of digital codes, but the recognition of an idea is a severe intellectual problem, the solution to which will probably never be exact. Nevertheless, this is the problem which must be attacked if accuracy is ever to be attained, or even approached, in using the text of information items as a basis for their recovery." 3/

Nevertheless, some of the work both in natural language text searching and in "probabilistic indexing" (where weights representing judgments as to degree of relevance of an indexing term to an item are used either in indexing or search), provide instructive insights into some of the problems of automatic indexing.

In the period 1958-1960, work at Ramo-Wooldridge resulted in the release or publication of provocative papers by Maron, Kuhns, and Ray on "probabilistic indexing" (1959 [398], 1960 [397]) and by Swanson on natural language text searching by computer (1960 [587, 582], 1963 [583]). Subsequent work along these lines has included further developments at Thompson Ramo-Wooldridge, the law statutes work at the Health Law Center at the University of Pittsburgh, and the experimental investigations of Eldridge and Dennis in a project jointly sponsored by the American Bar Foundation, IBM, and the Council on Library Resources.

1/ Doyle, 1959 [168], p. 2.

2/ Salton, 1962 [520], p. III-1 through III-2.

3/ Doyle, 1959 [165], p. 12.

6.4.1 Probabilistic Indexing - Maron, Kuhns, and Ray

The work in the area of "probabilistic indexing" involves, as in the case of Stiles' statistical association factors, an assumption that there should be machine means available for the automatic elaboration of search requests in order that relevant documents not indexed by the precise terms of these requests may be retrieved. Given that measures of "closenesses" and "distances" between similar documents can be obtained, probabilistic weighting factors between index terms assigned to documents may be made explicit. More generally, however, the notion of probabilistic indexing is based upon the assignment of weights that provide a numerical evaluation of the probable relevance of index terms to a particular document, and of the relative importance of the various terms used in a search request. Maron and Kuhns (1963 [397]) thus consider the following variables important in the formulation and following out of search strategies:

1. Input- both the terms of the request and the weights assigned to them.
2. A probabilistic matrix giving dissimilarity measures between documents, significance measures for index terms, and closeness measures between index terms.
3. A priori probability distribution data.
4. Output- a class of retrieved documents ranked in order of their "computed relevance numbers" and an indication of the number of documents involved in the class.
5. Search parameter controls, such as the number of documents desired.
6. Search prescription renegotiation involving amplification of the request by adding terms "close" to the ones in the original request and the selection of additional documents following distance criteria for the collection.[1/]

Experiments have been reported for 40 requests run against 110 articles taken from Science News Letter. Without search renegotiation, the "answer" document was retrieved in only 27 of the 40 tests. Three alternative methods of request elaboration were then tried. First, additional terms most strongly implied, statistically, by the terms in the request were used. Secondly, those terms were added which most strongly imply, again in a statistical sense, each of the given request terms. Thirdly, coefficients of association between index terms were used. Results are reported as follows:

"(1) Using the method of request elaboration via forward conditional probabilities between index tags, we retrieved the correct answer document in 32 cases out of the 40.

(2) Elaborating the requests via the inverse conditional probability heuristic, we retrieved the correct document in 33 of the 40 cases.

(3) Using the coefficient of association to obtain the elaborated request we obtained success in 33 cases of the 40.

[1/] Maron and Kuhns, 1960 [397], pp. 230-231.

"Thus we see that the automatic elaboration of a request does, in fact, catch relevant documents that were not retrieved by the original request." [1]

6.4.2 Natural Language Text Searching - Swanson

The work in automatic indexing and related research directed by Swanson at Ramo Wooldridge Corporation has included "indexing at the time of search" in natural language text searching, (1960 [582, 587], 1963 [583]), the previously mentioned studies of machine-like indexing by people (Montgomery and Swanson, 1962 [421]), and automatic assignment indexing using pre-selected lists of clue words, (Swanson, 1963 [580]). The last of these three major areas of investigation is the one of the greatest interest in this present study, but the earlier experiments in machine searching of natural language texts warrant some discussion. In his reports on this text searching project, Swanson has specifically claimed that the methods for transforming search questions can serve as the basis for an automatic indexing method. Thus:

> "...A technique for automatic indexing can be derived immediately from a text searching technique...it is necessary only to so organize the machine procedures that those operations of text reduction or reorganization common to all searches are performed only once and prior to searching in order to create directly an automatic indexing procedure." [2]

Swanson has also claimed that if automatic searching of full text is not feasible, then automatic indexing is not feasible, the one being prerequisite to the other. For example:

> "Clearly, if a computer technique for search and retrieval from the full text of a collection of documents cannot be developed, then it is unthinkable that matters could be improved by using the machine to operate on just part of the information (a 'condensed representation') -- that is, on an automatically produced index. This line of argument demonstrates persuasively that the development of techniques for automatic full-text search and retrieval is a prerequisite to automatic indexing. It is equally clear that a technique for automatic indexing can be derived immediately from a text-searching technique, and thus that the two processes involve conceptually equivalent problems." [3]

In the actual text searching experiments, a model "library" consisting of 100 short articles in the field of nuclear physics was set up in machine-usable form. These articles were also studied by subject specialists who rated the relevance of each paper to each of 50 questions, and assigned weighting factors representing the degree of judged relevance. A second group of people, who knew only that the papers were in the field of nuclear physics, then transformed the 50 questions into search prescriptions using three different methods. The first method for the development of the search instructions was to choose appropriate index entries from a subject heading list tailored to the contents of the sample library. Search was then made manually against a card catalog which recorded the results of manual indexing of the same 100 articles to the entries of this list.

---

[1] Ibid, p. 240.

[2] Swanson, 1960 [582], p. 6.

[3] Swanson, 1960 [587], p. 1100.

The second method of search prescription tested involved the specification of combinations of words and phrases likely to be found in any paper which would in fact be relevant to the search question. The third method involved modification of the second by the use of a thesaurus-type glossary which suggested various alternative terms. Both the latter two types of search instructions were fed to a computer program which carried out searches against the natural language text consisting of 250,000 words from the original articles.

The results were then evaluated in terms of ratings of relevance made by the physicists who had analyzed the papers. Retrieval effectiveness was not high: "...in no case did the average amount of relevant material ... retrieval (taken over 50 questions) exceed 42 per cent of that which was judged ... to be present in the library." [1] However, the results were indicative of the superiority of the machine methods to the manual catalog search.[2] For this library in particular, in the case of "source documents" (the articles from which the search questions were taken), only 38 percent of the relevant papers were located by the manual search, whereas 68 percent of the relevant items were retrieved by machine search of the text for specified words and phrases in various "and" and "or" combinations. Machine search based on search instructions that had been developed with the assistance of the thesaurus-glossary yielded 86 percent of the relevant source item documents.

6.4.3 Full Text Searching - Legal Literature

"The retriever of documents may be satisfied with a sample of descriptors that represent the contents; the fact retriever or the question answerer must often have access to every word in the text". [3] The objective of fact retrieval is a major goal in the experimentation that is being carried forward in the field of natural language text searching of legal material, especially the texts of statutes of the State and Federal Governments. The most extensive program to date is that of Horty and his colleagues at the University of Pittsburgh Health Law Center (1960 [277], 1961 [276, 309], 1962 [196, 278], 1963 [24, 280]).

Wilson at the Southwestern Legal Foundation is experimenting with a modified version of the Horty-Pittsburgh System for legal cases dealing with arbitration in five of the southwestern states.[1/] A joint American Bar Foundation--IBM research program has been established to explore both text searching without prior indexing and automatic indexing techniques (Eldridge and Dennis, 1962 [183], 1963 [182]).

---

[1] Swanson, 1960 [582], p. 25.

[2] Ibid, p. 1: "On the whole, retrieval effectiveness was rather poor, yet machine search of the text of the model library was significantly better than was human searching of the subject heading index."

[3] Simmons and McConlogue, 1962 [555], p. 3.

In the Horty-Pittsburgh System, approximately 6,000,000 words of text have been converted via Flexowriter to magnetic tape. An exclusion dictionary of 100 words is used to eliminate the most common words and a word-concordance is prepared, resulting in word-occurrence location indicia by position in sentence, paragraph and section of the statute. In searching, the user has available to him the alphabetized list of approximately 17,000 different words and it is up to him to think of the words and synonyms most likely to occur in statute sections likely to be the ones he seeks. Several search logics are available. One provides that at least one of a group of alternate words must appear; another requires that at least one from two or more groups must appear in the same sentence. Intra-sentence distance criteria are also utilized: "If the phrase 'born out of wedlock' is sought, the operator... requires that the word 'wedlock' appear in the same sentence, no more than three words after 'born'." [2/]

Obviously, for the same question the searcher would also have to specify synonymous words and phrases--"illegitimate children", "illegitimate births", "unwed mothers", "unmarried mothers", "illegitimacy", "bastardy", and so on. The reported success of the system is apparently due in large part to the ingenuity of the searchers in specifying the expressions and synonyms most likely to be used. Hughes comments as follows:

> "It should be noted that this system will be most efficient only when the users are thoroughly familiar with the linguistic style of the source material and search is made on words known to occur in the appropriate statutes". [3/]

6.5 Other Examples of Related Research in Linguistic Data Processing

Since, as Garvin has emphasized, "All areas of linguistic information processing are concerned with the treatment of the content, rather than merely the form, of documents composed in a natural language," [4/] much of the research in linguistic data processing is potentially applicable to both the development and the improvement of automatic indexing techniques. Thus developments in automatic content analysis, in psycholinguistics, in question-answering systems, may eventually find application to mechanized indexing systems.

[1/] Eldridge and Dennis, 1964 [182], p. 90; Wilson, 1962 [645].

[2/] Horty, 1962 [278], pp. 59-60.

[3/] Hughes, 1962 [284], p. IV-6 to IV-8.

In terms of our present concern, however, we shall select only a few examples. "By automatic content analysis is meant the use of computer programs to detect or select content themes in a sentence-by-sentence scanning of text or verbal protocols".[1] The interest of psychologists in machine techniques to assist in the analysis of linguistically-given materials, as in propaganda analysis, probably precedes at least in sophistication if not by date, that of documentalists or of machine specialists interested in library and information problems.[2]

The "General Inquirer" program developed by Stone et al,[3] is an example of question-answering techniques based upon selective extractions from natural language text. It involves the use of a master vocabulary consisting of words previously selected by an investigator as being likely to be content-indicative in a body of material to be processed, together with his pre-established indications of the categories he expects their occurrence should predict. It is to be noted that this is a custom-tailored set of categories and of clue-word lists associated with each, manually pre-established. Text is now processed in such way that each word is looked up and, if it appears in the master vocabulary, it is tagged with identifiers of the categories for which it is presumably predictive. A subsequent "Tag Tally" routine then counts the tag frequencies to determine for which categories the input material has high or low scores, and these in turn can be compared with expected norms.

This type of program has been applied to such varied materials as suicide notes, folk tales from different cultures, reports of field workers, recordings of group discussions as in supervisory-leadership training sessions, and protocols for various psychological tests.[4] Interesting variations developed by Jaffe and others[5] involve the use of non-verbal as well as verbal clues as content-indicators, specifically, time-sequence patterns recorded along with the words spoken in client-therapist sessions. At the meeting of the Association for Computational Linguistics and Machine Translation held in Denver, August, 1963, Jaffe reported findings indicative of positive correlation between the structure of temporal and lexical patterns in dialogue and suggested applications to automatic abstracting or indexing by the use of the time-sequence patterns as clues to high information-value areas.

---

[1] Ford, Jr., 1963 [498], p. 3.

[2] See, for example, Jaffe 1952 [297], Hart and Bach, 1959 [256], Pool, 1959 [475], the latter covering the proceedings of a conference held in 1955.

[3] Stone and Hunt, 1963 [576]; Stone et al, 1962 [575].

[4] See Ford, 1963 [498], p. 8.

[5] See for example, Cassotta, et al, 1964 [104]; Jaffe, [294] to [297].

Hughes provides, as of September, 1962 ([284]), a critical review of several experimental and proposed question-answering systems using natural language statements and natural language queries, including "BASEBALL", [1/] "SAD SAM" [2/] and the "Proto Synthex" investigations of System Development Corporation. [3/] Later developments on the Synthex (synthesis of complex verbal material) project at SDC have included a variation on a natural language text searching program where ordinary text input is run against an exclusion list and a table is set up to tally the substantive words remaining. Words with the same roots or previously having been identified as synonymous are cross-referenced. A complete index results, with document location identifier tags for the word occurrences down to the single sentence level. This index can be used subsequently to locate regions of text (volume, chapter, paragraph, and sentence) where answers responsive to input questions are likely to be found.

It is proposed that the Synthex system eventually should incorporate analyses of syntactic and semantic relationships in the linguistic expressions of both queries and text. Of future interest in the extension of such considerations to automatic indexing and abstracting are the following comments:

> "The results of several early experiments within the project, coupled with the findings of other language researchers, led to the following conclusions about meaning and grammatical structure in English text:
>
> 1. The degree of synonymity in meaning between any two English words can be measured quantitatively with a synonym dictionary and relatively simple scoring procedures.
>
> 2. The difference in meaning between two sentences of identical syntactic structure can be expressed quantitatively as a function of synonymity of their words..." [4/]

It is also of interest to note that although the "indexer" program of the Synthex system provides cross-referencing between, for example, "whales" and "whaling" or "England" and "Great Britain", the investigators admit that: "naturally it falls short of such complicated cross-referencing as 'mouse-animal' 'Jones person' and other concept recognitions." [5/] However, concept recognitions based upon both a priori and

---

[1/] See also Green et al, 1961 [238].

[2/] See also Lindsay, 1960 [363].

[3/] See also Klein and Simmons, 1961 [325]; Simmons et al [552] to [555].

[4/] System Development Corporation, 1962 [590].

[5/] Simmons and McConlogue, 1962 [555], p. 70.

a posteriori associations are at least foreshadowed in a small-scale model of attribute-words and proper names, together with prespecified relationships between them;[1] in Olney's recent work at SDC exploring the possibilities for use of cognitive concepts as bases for establishing association between documents,[2] and by Kochen's work on machine' inference and concept processing.[3]

A final example of potentially related research in the area of content analysis is therefore the work of Kochen, Abraham, Wong and others at IBM's Thomas J. Watson Laboratories (1962 [329]). While concerned principally with adaptive organization and processing of stored factual statements and the possibilities for machine formulation of "hypotheses" about these and additional facts, some consideration has been given to sampling procedures applicable to determination of similarity which might be used for document clustering and to the possibilities for dynamic clustering for retrieval based upon a specific individual query.[4] In the proposed AMNIP (Adaptive Man-machine Non-Arithmetical Information Processing) system„ there is no attempt at either automatic indexing or automatic abstracting.[5] Instead, formal statements are made about named "things" and their attributes. The sharing of common attributes then serves as a basis for relating items which are similar and for grouping them together in the system memory. It is assumed that the organization of the stored statements changes dynamically with new data inputs and user feedback in question-answering routines.

Where the named items are names of documents or of index terms, a number of documentation applications can be considered. Where the items are document names and the formal predicate is "cites", the system provides a procedure for production and use of citation indexes.[6] Where the items are index terms or subject headings and the predicates are "is used synonymously with" or "is subsumed under", machine construction of a growing thesaurus based on use is suggested.[7] The common attribute matching program, applied to logical similarities of texts related as by having various assigned descriptors or citations in common, might provide a basis for generating document surrogates by representing each text in a related group of texts with the words or sentences these texts have in common. 1/

---

[1] See Stevens, 1960 [568]; see also Herner, 1962 [266], p. 5.

[2] See Borko, 1962 [75], p. 5: "Instead of defining meaning in terms of synonyms... it is defined in terms of the entities referred to by the word in context. A chair is thus described as belonging to a class defined by a given list of properties... Analysis yields an interpretation of the sentence as an assertion that certain relationships hold between the specified referent classes. The cognitive content of the sentence is a function of this assertion plus the information about these referent classes which has previously been stored in memory."

[3] Kochen et al, 1962 [329].

[4] Ibid, Appendix by C. T. Abraham, pp. 20-65.

[5] Kochen et al, 1962 [328], p. 45.

[6] Ibid, p. 37.

[7] Ibid, p. 37.

In the case of man-machine interaction during search, it is suggested that the user should indicate the names of selected documentary items which are of particular interest, then:

> "The machine forms an 'hypothesis' about the subset of articles likely to be of interest. It does this by examining all recorded statements common to the ones selected but not to the rejected ones. The weight of different attributes and degree of interest is taken into account. The machine may display this hypothesis or another random sample of titles consistent with it, or both." 2/

6.6 Machine Assistance in Translations of Subject Content Indications to Special Search and Retrieval Language

There are, also, in the areas of directly and indirectly related research, certain programs of research, development, and experimentation which include investigations of possibilities for using machines to assist in the "translation" of textual languages into special intermediate or "documentary" languages. Doyle's use of the inclusion list principle to extract specified content-indicative words and to encode them in his "bigram" index was an early but relatively trivial example. 3/ The work of Williams and her associates, at Itek and elsewhere, 4/ has involved the objectives of determining which of the subject-revealing implications of titles, abstracts and, if necessary, full text, are susceptible to machine detection and manipulation such that the implied as well as the explicit assertions made in a document may be incorporated in a formalized language for retrieval.

While Williams, Barnes, Cardin and Levy, and others, have so far approached such tasks primarily from the standpoint of human analytic judgments, Coyaud (1963 [143]) has discussed at least preliminary work looking toward the automation of the analysis of natural language texts for purposes of encoding and organization of the terms and relationships to be used in the "documentation language" known as "SYNTOL" (Syntagmatic Organization of Language), this work has used a corpus based on bibliographic abstracts from the Bulletin signalétique of the Centre National de la Recherche Scientifique, Psychophysiology Section, for the period 1958-1960. Notwithstanding such difficulties as determining rules for proper subdivisions of text, reduction of synonyms, resolution of lexical and syntactic ambiguities, and the fact that some words are always,

---

1/ Kochen et al, 1962 [329], p. 2.

2/ Kochen et al, 1962 [328], p. 7.

3/ Doyle, 1959 [168]. See also p. 123 of this report.

4/ See, for example, T. M. Williams, R. F. Barnes, Jr., J. W. Kuipers, various references.

but some never, used in SYNTOL itself, he reports that both substantives and textual expressions indicative of certain specific SYNTOL relations can be unambiguously identified. Contextual clues are used: for example, if the word "homme" occurs it is translated as "sexe masculin" if "femme" also occurs, as "etre humain" if "animal" is also mentioned, and as "sujet experimental" otherwise.

Melton and her associates at the Center for Documentation and Communication Research, Western Reserve, have also been investigating machine processing of input text with a view to the automatic selection and manipulation of clue words and relationships between them for information retrieval purposes. Their material consists of abstracts from the metallurgical section of Chemical Abstracts. From sample abstracts, a lexicon is developed which involves classification of words into those that are significant from a metallurgical point of view;those that name materials, compounds, environments; those denoting processes; those denoting characteristics of materials; prepositions; those which will not operate in the analysis of the text, and the like.

On the basis of analysis of a number of sentences from the sample text, rules for combination and selection of specified words in specified relationships can be set up. These rules are designed to identify sentence types which:

(1) Describe performance of a process on a material.

(2) Discuss a material in terms of properties, components, form, or environment.

(3) Describe a process without reference to specific materials.

(4) Discuss metallurgical properties without reference to specific materials.

(5) Discuss two or more materials, properties or processes.

(6) Describe a causal relationship between two properties.

(7) Give a comparison of materials.

(8) Contain no words of interest in the system.

Computer programs to explore the possibilities for automatic analyses of the kind developed manually for the sample abstracts will be written with the objective of finding an effective compromise between mere word identification and total linguistic analysis. Melton says:

> "If one considers this method of analysis from the point of view of the linguist, he can immediately describe many grammatical constructions, which will prevent the meaningful reduction of these sentences. It is not known at this time how often such sentences will appear in the corpus of this investigation. Nor is it known how adversely such failure would affect the retrieval of the information in these sentences. The answers to these questions will be available only after a large sample has been analyzed and put to an extensive retrieval test. At its most successful the project will achieve an automatic processing of metallurgical text which will permit retrieval of the type of information which can be stated in its own terms with a tolerable amount of inappropriate selections. Should this goal be unattainable, the project will have generated a file of abstracts automatically searchable on the word level

> or somewhat beyond. For the benefit of other research, it will also have produced tapes of the true text of a large sample of natural-language abstracts and a lexicon containing all the words of a corpus of current scientific literature." 1/

6.7 Example of a Proposed Indexing-System Utilizing Related Research Techniques

In addition to the automatic assignment indexing and automatic classification techniques for which experimental results have been reported, several other techniques and programs have been proposed. One is the joint American Bar Association-IBM research program (Eldridge and Dennis, 1963 [182]), for which discussion has been deferred because of its proposed use of several of the research techniques covered previously in this section. The experimental corpus will consist of the full text of approximately 5,000 legal case reports taken chronologically from the Northeastern Reporter. Approximately half of this material will be processed to obtain word frequency counts. The frequencies will then be used to prepare for each different word an estimate of the skewness of its distribution in the collection. The investigators will then personally inspect the word list as ordered by skewness to divide it into "non-informing" (Type I words, or an exclusion list) and "informing" (Type II words, or an inclusion list) at some appropriate cutting point. Then, for each document, a list will be prepared of its "informing" (Type II) words, maintaining order within the document. For each pair of such words, statistical association factors will be computed. Eldridge and Dennis describe other aspects of their proposed technique, in part, as follows:

> "For each document in the body of 2,500 cases, a list will be prepared of its Type II words, maintaining their original order within the document ... For each Type II word an 'association factor' will be calculated for every other Type II word with which it appears in any one document by compiling the probability that Word A would appear this close to Word B this number of tries over the entire file, if the Type II words were distributed at random. (This amounts to borrowing Stiles' idea of the association factor, but implementing it with a numerical method which takes into account nearness of the words within the document as well as the fact that they both occur in the same document.) Since the factors are probabilities, they will be numbers between zero and one ... These numbers will be used to estimate the distances between words in index-word space.
>
> "The next step is to construct from the information about distances between pairs of words an index-word space in which every word is at the correct (or approximately correct) distance from every other word in the system with which it exhibits association. The result of this operation can be visualized schematically as a sort of grid in which every word can be placed in its appropriate position by assigning it a set of coordinates."

[1/] Melton, et al 1963 [414], pp. 14-15.

> "Indexing of the remaining cases in the experiment will be performed by machine from full text, using the Type I list of discard words and the Type II list to prepare an analysis of the frequencies related to index-word space. Instead of selecting specific words as indexing terms, concepts will be selected (statistically) as volumes in index-word space. A rough physical analogy to this process would be to toss pennies at the previously mentioned grid so that, for every Type II word in the source document, a penny lands at its proper slot on the grid. Where the pennies heap up in a pile, you have a concept."
>
> "Searching will be carried out essentially by indexing a question presented narratively, determining the concept volumes that represent the question, and searching those volumes in document space for the relevant document numbers. Since the 'edges' of the concept volumes are determined statistically, output can be listed in order of probable relevance; as an option the question could be accompanied by a request that 'at least 100 references be supplied', in which case the concept boundaries would be adjusted to provide that number." 1/

It will thus be noted that the proposed indexing and search program begins on a derivative basis to establish for one-half the experimental material the significant words, next combines word frequency with significant word distance data to derive probabilistic association factors between words, then develops clusters, and finally indexes the items in terms of the clusters rather than words so as to provide assignment rather than extraction of index terms.

7. PROBLEMS OF EVALUATION

We have noted, in the introduction to this report, that several fundamental and highly controversial questions can be raised with respect to the feasibility and evaluation of any automatic indexing scheme and with respect to the evaluation of any indexing systems whatsoever. Yet if automatic indexing procedures are to be based upon previous human indexing or if their results are to be compared with human results, then the questions of the quality, the reliability and the consistency of human indexing are crucial ones indeed. Thus, Solomonoff warns:

> "The finding of exact languages for retrieval is also made less likely, in view of the fact that the categorizations of documents that are presented to the machine as a training sequence will not be performed altogether consistently by the human cataloger." 2/

Montgomery and Swanson ask whether human indexers are in fact self-consistent and consistent with each other, and they suggest:

1/ Eldridge and Dennis, 1963 [182], pp. 97-99.

2/ Solomonoff, 1959 [562], pp. 9-10.

"If the answer turns out to be 'no', we might reasonably conclude that the only reliable and effective kind of human indexing is that which is already machine-like in nature." [1]

With a few noteworthy exceptions, there has been very little serious investigation of these problems and there is very little comparative data.

O'Connor has been making a series of studies, with considerable emphasis upon how one might measure the products of machine indexing and how one might derive machine rules for automatic index ing from systematic review of documents indexed by people. Cleverdon and his associates at the ASLIB Cranfield project have extensively tested several different indexing procedures. Painter, MacMillan and Welt, Slamecka and Zunde, and others report findings on intra-indexer, and inter-indexer consistency -- unfortunately, on the basis of quite small samples. Various alternate approaches to the evaluation of automatic indexing results have been considered by Borko, Doyle, Swanson, Savage, Giuliano, and others. In addition, some data bearing on these questions have been reported in connection with analyses of selective dissemination (SDI) systems. Some data from other sources, such as studies of user preferences with respect to various reference and search tools, is also pertinent.

The most generally accepted criterion for appraising the effectiveness of indexing is that of retrieval effectiveness. But, in general, this is merely the substitution of one intangible for another, entailing a string of as yet unanswerable or at least unresolved questions.[2] Retrieval of what, for whom, and when? How can effectiveness be measured except by the elusive question of relevance judgments? How can human judgments of relevance and value be measured and quantified?

We shall try to distinguish here, insofar as possible, between the core problems that make the evaluation of indexing as such an extremely difficult task, the available data on human indexer reliability, and the possible advantages and disadvantages of automatic indexing techniques.

---

[1] Montgomery and Swanson, 1962 [421], p. 366.

[2] Compare Swanson, 1960 [582], pp. 2-3: "The performance of retrieval experiments when relevance judgments per se cannot be consistently assessed by human judgment would seem to represent overly vigorous pursuit of a solution before identifying the problem." Similarly, see Black, 1963 [64], p. 14: "Finally, when one is faced with an existing collection of indexed materials, how does one assess the effectiveness of any retrieval system? Suppose that one receives 20 documents as a result of a query to the system. Suppose further that all 20 documents are quite pertinent to the topic of interest. Is there any way to assess the amount of pertinent information still unretrieved from the file? Or is there any way of learning whether the retrieved information is more pertinent than the unretrieved information ? The answer is 'No!' -- the use of any retrieval system is, then, an act of faith in the quality of indexing."

7.1 Core Problems

First and foremost of the core problems implicit in the question of evaluation of any indexing scheme, whether applied by man, machine, or man-machine combinations, are those of interpersonal communication itself, which in turn relate to fundamental problems of epistemology. These are, first, the problems of language as a means of communicating perceptions, apperceptions of relationships between present observations and prior experience, and value judgments based thereon, and, secondly, even more fundamentally, the question and the veridicality of language representations of real transactions and events. Serious investigators in the field, including many who have themselves contributed to automatic indexing techniques, have made such typical acknowledgments of the difficulties as the following:

> "The imprecision connected with discussion of retrieval effectiveness and of relevance is not due to lack of understanding of the relatively straightforward retrieval processes, but is due to our lack of basic understanding about language, meaning and human communication itself." 1/

> "Fundamentally, the study of inquiry procedures is a problem in the general psychology of cognitive functioning. Relevant problems concern the way problems are recognized and formulated into questions, the way a search plan is developed to find answers to questions, and finally, the way it is decided whether or not a possible answer matches the specifications of a question." 2/

A second core problem is the heterogeneous and somewhat arbitrary development of natural languages themselves. It is much the same fundamental problem whether men or machines are to read text and determine the "meaning" (at least, in the sense of communication intent) of messages expressed in a natural language. However, the problems are aggravated if men themselves must know enough about language and its conveyances of message content to specify precisely to a machine what it is to look for and to use.

Salton enumerates some of these difficulties as follows:

> "No well-defined set of rules is known by which the individual words in the language are combined into meaningful word groups or sentences. Specifically, the correct identification of the meaning of word groups depends at least in part on the proper recognition of syntactic and semantic ambiguities, on the correct interpretation of homographs, on the recognition of semantic equivalences, on the detection of word relations, and on a general awareness of the background and environment of a given utterance." 3/

1/ Giuliano, 1963 [230], p . 6.

2/ Stone, 1962 [576], p. 1.

3/ Salton, 1963 [519], p. I-2.

Similarly, Baxendale states:

> "We are confronted with difficulties which arise from the multiple ways in which words and sentences are put together to convey meanings and shades of meaning -- i. e., to represent ideas and concepts. Research into this problem -- drawing upon psychological and logical analysis -- is scarcely begun." [1]

A third core problem is the proper choice of appropriate selection criteria if condensed representations of document content must be used for scanning, search, and relevance decisions. Swanson suggests that the price paid for brevity of representation so that searching operations can be efficiently managed is the loss of at least some, perhaps most, of the information in a collection or library. He notes also that:

> "It is another obvious but seldom remarked fact that the extent of such information loss for existing libraries is not only unknown but has never defined in measurable terms." [2]

This loss is lived with, today, in many practical situations involving abstracts, index term sets, selective-dissemination notices, and even mere author-title listings in announcement bulletins or search output products from either manual or machine searches. Yet the sheer increase in volume of the total number of items to be covered and of the number of items potentially responsive even to a single individual's interests has severely stretched any individual's capacity to scan or skim, much less read, the presumably pertinent material -- documents themselves, abstracts of other documents, listings of documents available -- already accumulating on his desk.

Condensation, reductive representation, becomes more and more imperative. Concurrently, while conventional tools may be lived with, after a fashion, the substitution of machine-compiled or machine-produced alternatives, even though they give the same information in the same volume, number of pages to be scanned, may because of such things as inferiorities of page and line formatting, size of type on the page, limitation of typography to upper case and a few other symbols, make the problem of how adequate the user judges the selection and condensation to be, that much worse.

A fourth problem in evaluation, therefore, is the question of whether or not the benefit to users is worth the cost. For example, despite the arguments for concept rather than word indexing, for assignment of labels rather than mere extraction of a few words used by the author himself, at least some data on the use made by scientists of various sources of information on material which might be of interest to them suggests that subject indexes are not the most important source, nor even a major source. Herner found, for example, that only about 16 percent of his respondents reported use of indexes and abstracts as primary tools in literature searches. He reports, for the use of tools in becoming aware of current sources of information, 477 of 3832 responses indicating the use of indexing and abstracting publications as against 486 using footnotes or other cited references, [1] 291 using library acquisition lists, and 212 using separate bibliographies (Herner, 1958 [265]).

---

[1] Baxendale, 1962 [42], p. 68.

[2] Swanson, 1960 [582], pp. 5-6.

These data, and similar findings of Fishendon that 17 percent of scientists queried considered the scanning of titles in accession lists and announcement bulletins a principal means to find information of interest, [2] suggest that KWIC type indexes may be adequate for many purposes. On the other hand, the KWIC index to the U. S. Government Research Reports made available to the public on an experimental basis through the Office ot Technical Services was discontinued after a year of subsidized operation because too few of the users indicated willingness to pay a fee in order to have the service continued on a subscription basis.

The evaluational problem here involves the lack of information on indexing costs, the relatively few quantitative and objectively validated studies that have been made of user needs, the question of whether what the user says he does or wants is what he really wants or does, and the matter of defining "interest" for different users with differing purposes and requirements. The concept of "interest" is taken to mean the motivations of a particular user or group of users at a particular time, while the equally imprecise notion of "relevance" refers to the value judgments made by the user as to the relation of an item to his query or interest.

A final core problem, then, is that of the question of relevancy itself, involving recognition that "relevancy is a comparative rather than a qualitative concept ... (and) ... that a document of little relevancy in the eyes of X might well be highly relevant in the eyes of Y." [3] Mooers states, similarly, that:

> "There is no absolute 'Relevance' of a document. It depends upon the person and his background, the work and the date. What is not relevant today may be relevant tomorrow." [4]

Good discusses various possible measures of 'relevance' - logical measures, frequency measures, references to, citations of, interest measures, linguistic measures, [5]

---

[1] Note that Herner's data and those of Glass and Norwood, 1958 [232], reporting 6.9 percent use of cross-citations in another paper as the method of learning of important work as against 1.2 percent using an indexing service, appear to re-enforce the claims of those who advocate citation indexing.

[2] Fishenden, 1958 [197], p. 163.

[3] Bar-Hillel, 1959 [33], p. 4-8.4.

[4] Mooers, 1963 [423], p. 2.

[5] Good, 1958 [234], pp. 7-9.

but except for the obvious statistical criteria, the problems of how to measure relevancy remain largely unresolved.

At least some data on the variability of relevance judgments is available in reports of the performance of an SDI (Selective Dissemination of Information) system. In such systems, the indexing terms or tags assigned to a new item are compared with a file of "user-profiles" that is, with a pre-prepared listing of terms or topics in which a particular user is interested. Where the term-profile of a new item matches that of a user, a notification of the acquisition of that item is sent to him. Barnes and Resnick report tests of such a system in which pseudo-notifications selected randomly were included with those produced from the matching procedure. Account was kept of which notices were regarded by the users as meeting their interests and which were not. They found that 58.1 percent of the non-random notifications were regarded as relevant, but that so also were 26.8 percent of the random ones. 1/

Katter comments on findings that the intersubjective agreement of typical users with respect to value judgments of condensed representations of text is low. He suggests:

> **"One source of this low intersubjective agreement among users may be that it is often not clear what is intended by the words relevant and representative. Considerations such as the validity of the material, its usefulness, stylistic qualities, understandability, conceptual preferability, etc., can all enter their judgments in unknown amounts." 2/**

Corroborating evidence is available from other sources. Swanson, in his tests of a natural language text searching technique, had first used subject matter specialists to rate the relevance of each of the text documents to each of 50 questions. Two individuals rated each item, and if they disagreed significantly, a third person was asked to reconcile the difference. In spite of this, 8 percent of the cases of failure to retrieve "relevant" documents were ascribed to incorrect initial judgments of relevance, and 15 percent of the presumably "irrelevant" documents were finally judged to be relevant after all (Swanson, 1961 [586]). In Swanson's words: "The question of formulating criteria for judging the relevance of any document to the motive, purpose, or intent which underlies a request for information is profound and lies at the heart of the matter." 3/

1/ Barnes and Resnick, 1963 [36], p. 2.

2/ Katter, 1963 [308], p. 24.

3/ Swanson, 1960 [587], p. 1099.

7.2 Bases and Criteria for Evaluation of Automatic Indexing Procedures

What should the bases be for the evaluation of existing or proposed indexing systems that rely, to a greater or lesser extent, on machine generation of the indexing or classificatory labels? Since the evaluation of quality of indexing per se raises such fundamental and elusive questions, can these questions be begged for the case of automatic indexing as they are in fact for almost all manual systems? If so, the obvious bases are those of time, cost, availability of alternative possibilities, and customer acceptance. Here again we are faced with a dearth of objective data, even for the intercomparison of any two manual systems.

In the two years preceding the ICSI Conference, the Program Committee openly solicited papers that would provide comparative data for operating information systems and that would develop and discuss criteria for the comparison of systems. [1] Nevertheless, of the papers received only two were responsive to this invitation: the special case of comparing the conventional file against the inverted file approach to the searching of chemical structure data (Miller et al, 1959 [419]), and an early report by Cleverdon on the ASLIB Cranfield project for the intercomparison of indexing systems, under a grant from the National Science Foundation (1959 [126]).

There had been an earlier comparative experiment, generally conceded to be the first of its kind, [2] in which 98 search requests were run by ASTIA personnel using a conventional catalog and by personnel of Documentation Inc., using a coordinated uniterm index. Warheit says:

> "Unfortunately, the conditions of the test were very poorly designed so that, in the final analysis, each group was the sole judge both of the scope of the original request and of the adequacy of the bibliographies produced. The resulting claims are of course contradictory." [3]

---

[1] See "Proposed Scope of Area 4," Proceedings, ICSI, 1959 [481], pp. 665-669.

[2] Compare, for example, Gull, 1956 [246], p. 329: "When one considers that a fairly thorough search of the literature indicates that this comparison of two reference systems is the first undertaken so far, it is not surprising that the results reveal clerical errors and an incomplete design of the test."

[3] Warheit, 1956 [631], p. 274.

However, some of the findings are pertinent to our present questions of evaluation. Thus, of 492 items selected by Documentation, Inc., that ASTIA considered pertinent but had not selected, 98 were missed by them although the proper subject heading was searched and the catalog card had adequate selection clues, 89 were missed because not all applicable subject headings were searched, 21 were missed because the original subject heading assignments had been inadequate, 7 were missed because neither title nor abstract provided indication that the report itself was pertinent to the request, and 102 were missed "because the subject heading did not occur to the searcher or because there were so many cards under the subject heading that the searcher was discouraged". [1] Similarly, Gull reports, of 318 items selected by ASTIA that Documentation, Inc. personnel considered relevant but had not themselves selected, 97 were missed because the searcher did not consult the proper terms.

7.2.1 The Cranfield Project

The inauguration of the Cranfield project is itself indicative of a prior lack of objective standards as applied to the measurement of effectiveness of information indexing, selection and retrieval systems. [2] Beginning in 1957, and still continuing with respect to individual indexing devices such as synonym controls and role indicators, this work has attempted to compare different indexing systems (e.g., UDC, Uniterm, etc.) under different indexing conditions (e.g., type of training of indexer, length of time allowed to index) against proposed measures of "retrieval effectiveness". These measures are, respectively, the recall ratio, or the percentage of relevant documents retrieved as against the total number of relevant documents known to be in the collection, and the relevance ratio, or the percentage of relevant documents among those actually retrieved.

In the first Cranfield tests, on 18,000 documents, it is reported that the recall ratio ranged between 75 and 85 percent for all four indexing systems. [3] These results are

---

[1] **Gull, 1956 [246], p. 329.**

[2] **Compare, for example, Randall, 1962 [492], pp. 380-381: "Prior to 1957, the proponents of the various indexing and classification schemes, the universal decimal system, the alphabetic subject heading, the Uniterm system and faceted classification touted their own system on the bases of subjective evaluation and theoretical investigations. There were many claims and much supposition about the relative merits and benefits ... but there was no body of data from which an objective evaluation could be made...Many observers believe that the Cranfield study constitutes the most important work done in the field of cataloging in recent times."**

[3] **Cleverdon, et al, 1964 [130], p. 87.**

rather better than reported by others [1] and have been subjected to specific criticisms although these first tests were limited to the recall of the source documents on which the test questions were based. For non-source documents there would of course also be questions relating to the core problem of how relevance is to be judged. Thus Markus says:

> "Despite investigations by Cleverdon in England, and by many others, there is today no generally accepted method of comparing the effectiveness of different types of indexes. The needs of index users vary so greatly that even the most carefully planned tests of retrieval efficiency can be challenged." [2]

Notwithstanding such criticisms, however, and in spite of the fact that the Cranfield tests have so far been directed principally to indexing systems applied manually, certain findings and conclusions reached by Cleverdon and his associates are pertinent to the questions of evaluating automatic indexing procedures. Examples are:

> "The fact is that no indexing sleight of hand, no indexing skill, can produce a system in which a figure for recall can be improved substantially without weakening the over-all relevance, i.e., the number of documents that are really relevant compared with the total number retrieved.
>
> "The majority of the failures (60 percent) were due to inadequacies and inaccuracies (carelessness rather than lack of knowledge) in the indexing process. However, supplementary tests, in which the staff of outside organizations carried out the indexing revealed that the Cranfield indexers were achieving a standard above average. This seems to indicate a certain inevitability of human weakness and error in the indexing process and lends some support to the many current research projects that are investigating the feasibility of automatic indexing." [3]

7.2.2 O'Connor's Investigations

As O'Connor has cogently observed on a number of occasions, the question of whether or not automatic indexing is possible is not the real question. Rather, the problem is whether or not indexing by machine is capable of producing results that are "good enough" for retrieval purposes, raising in its turn the still more basic question of how "good retrieval" can be evaluated. His own approach in detailed investigations has been to study an existing system (e.g., using Merck, Sharp and Dohme data) with respect to indexing terms such as "penicillin," "toxicity," and "mode of action." He then attempts to define various possible machine assignment rules, and then to determine the probable over-and-under assignments that would result from the application of these rules.

---

[1] See, for example, Johnson 1962 [300], p. 90: "The amount of meaningful information that can be retrieved is too small. There are few available studies on this subject. But these seem to indicate that, under some indexing schemes, meaningful retrieval can run as low as 10 and 15 percent and that the most that can be optimized for any of them, even under highly motivated conditions, is around 70 percent."

[2] Markus, 1963 [394], p. 16. See also Kochen, 1963 [327], p. 12: "The outstanding large-scale and realistic experimental work is that of Cleverdon. Unfortunately, his results are not very decisive."

[3] Cleverdon et al, 1964 [130], pp. 86-87.

Typical results pertinent to both questions of word-indexing evaluation and of inter-indexer consistency showed that for 23 documents indexed under the term "toxicity," 11 did not contain the stem "toxi..." at all; that 17 items indexed under "penicillin" contained the word at least once; that none of 34 randomly selected documents not indexed under "penicillin" contained the word, but that 7 of 28 items not so indexed but selected as probable candidates from title and other clues did contain the word. (O'Connor, 1961 [447])

Typical suggestions, comments, and conclusions made by O'Connor include the following:

> "It might be required that the mechanized indexing permit as good (or no worse) retrieval as existing human indexing, because it is desired to free the subject-skilled indexing personnel for other work. Or poorer retrieval (than possible with human indexing such as is presently done of comparable material) might be accepted from computer indexing, because poorer retrieval is better than none and there is a shortage of subject-skilled people to do the additional indexing." [1]
>
> "Such considerations as the following are relevant. Over-assigning can increase input costs and storage (to an extent dependent on the storage system), but mechanizing indexing might be worth the cost. Over-assigning might also increase the number of irrelevant documents retrieved, but the increase might be insignificant." [2]
>
> "...Suppose terms A, B, and C each correctly characterize five percent of a ten thousand document collection, each term is overassigned to another five percent, and over-assignment of each term occurs independently of the correct assigning and over-assigning of the others. Then about nine documents will be extra for the search question A & B & C." [3]
>
> "The question of permitting some under-assigning, that is, the computer failing to assign [a term] T to some document which should have it, is more delicate. Human indexers sometimes underassign. If we knew the rate of ounderassigning by human indexers for a term T, we might consider allowing the computer a similar rate. However, some cases of underassigning might be more important than others and if the computer made more important mistakes than the human indexers, retrieval might not be 'good enough'." [4]

---

[1] O'Connor, 1960 [444], p. 3.

[2] O'Connor, 1961 [448], p. 199.

[3] O'Connor, 1960 [444], p. 6.

[4] Ibid, pp. 6-7.

Other typical points made by O'Connor include the possibilities that the use of automatic indexing techniques might free trained technical people for other work, that it might permit more indexing than is now possible with available resources, that it might cost less, and that it might produce a better or more consistent indexing product.[1/] With respect to the latter point, however, he points out that greater consistency might not in itself be a virtue, since the product although generated more consistently might be relatively worthless by comparison with the inconsistent human product. [2/] Especially pertinent to the question of judgment factors in evaluation was a comparison of the most frequent words selected by the Luhn "auto-encoding" technique as applied to an ICSI paper against a quasi-random word list for the same paper produced by selecting the last non-common word on every page, and the first such word on every second page. He remarks:

> "The important point of this quasi-random list for my present purposes is to emphasize that first impressions might not be at all a good way of judging the adequacy of an index set." [3/]

7.2.3 Questions of Comparative Costs

The paucity of objective data on the effectiveness of indexing systems generally extends to even such obvious questions as costs of indexing and time required to index. These very questions might, in fact, be decisive with respect to choice between manual and machine systems. It has been estimated by some that the costs of manual subject indexing amount to close to 75 percent of the costs of operating an information selection and retrieval system, [4/] yet very little actual data on costs has been reported in the literature. [5/] Exceptions are, for the most part, limited to rather special cases, such as the following examples:

1. A total cost of less than $30,000 is reported for a 10,000 document collection at Aeronutronic. Four man-years of effort were required. On average, 12.6 access points were provided per document, of which 9.2 were subject-indicating descriptors chosen, with some modifications, from the second Edition of the ASTIA Thesaurus. "This favorable figure was possible because an adequate ready-made thesaurus of indexing terms was available and because the 'peek-a-boo' type equipment used was much

---

[1/] O'Connor, 1962 [447], p. 267.

[2/] O'Connor, 1963 [443], p. 16.

[3/] O'Connor, 1962 [447], p. 270.

[4/] O'Connor, 1963 [442], p. 1.

[5/] See, for example, A. D. Little, Inc. (1963 [23], p. 5): "Performance and cost data on existing large documentation systems are surprisingly sparse, and cost data have rarely included adequate overhead and depreciation accounting."

less expensive than most other devices offering comparable speed of operation and search logic possibilities." 1/

2. **"The experience of libraries that have gone through indexing using links and role indicators and careful editing shows that indexing takes about one-half hour per document (or $4.00) and costs an additional $1.00 for routine processing." 2/**

3. **In an investigation of the comparative merits of manual indexing of 2,000 documents using the UDC classification system as against a KWIC index, Black gives the figure of approximately $1400 for the UDC case compared to about $600 for an in-house computer operation to produce KWIC listings, and somewhat more for a KWIC index compiled by a service bureau. 3/**

Time required to index, which directly involves cost, is reported by Cleverdon to vary widely:

> **"Few reliable figures have been given for current practices, although a particularly high figure is the 1 1/2 hours average quoted for indexing reports for the catalogue of aerodynamic data prepared by the Nationaal lluchtvaart laboratorium in Holland. It appears from personal discussions that an average of 20 minutes for a general collection of technical reports is the top limit, and this has been taken as the maximum indexing time to be used in the project." 4/**

Insofar as such meagre data is indicative, there does not appear to be any particular cost-advantage for machine-compiled and machine-generated indexing other than the title-only KWIC indexes. Thus, Olmer and Rich report, in part:

> **"The program ... lends itself to a variety of applications. One of these ... is estimated to cost roughly $4.00 per document for cataloguing, putting on tape, printing and making any necessary corrections." 5/**

This is for a case where the indexing (cataloging) is done manually.

For a specific proposed automatic indexing system, employing a modified version of the Luhn word-frequency counting selection principle, Gallagher and Toomey report that:

1/ Linder, 1963 [361], p. 147.

2/ Lockheed Aircraft Corp., 1959 [369], p. 93.

3/ Black, 1962 [65], p. 318.

4/ Cleverdon, 1959 [126], p. 690.

5/ Olmer and Rich, 1963 [454], p. 182.

> "For the documents in our system, we estimate that processing time will be about 20 seconds per thousand words ... The cost is approximately $3.50 per minute when averaged between prime and extra shift." 1/

This means that the cost of processing a 3,000-word document would be $3.50 , exclusive of the costs of keypunching the input text which, conservatively estimated, costs not less than 1-2 cents per word. 2/ Swanson similarly assumes either that machine-usable text is already available or that editing and keystroking efforts are separate costs in arriving at an estimate of $1.00 per item for automatic indexing. 3/

These quantitative estimates bear out the more subjective conclusions of such investigators as Bar-Hillel, O'Connor, and others. Examples are:

> "It is very likely that manual uniterm indexing by cheap clerical labor will still, on the average, be qualitatively superior to any kind of automatic indexing, and it is very unlikely that the cost of automatic indexing will ever be less than this kind of manual uniterm indexing, unless the automatic indexing is to be of such low quality as to totally defeat its purpose." 4/

> "Most of these techniques require that the full texts of documents be in machine readable form. At present this usually requires keypunching which is much more expensive than a specialist's indexing efforts." 5/

1/ Gallagher and Toomey, 1963 [205], p. 52.

2/ 'Compare, for example, Ray, 1961 [496], p. 55; Swanson, 1962 [584], p. 470: The cost is roughly one or two cents per word which by standards of what is normally spent for even the most thorough indexing and cataloging, is exorbitant." Mersel and Smith report 1964 [415], p. 10A) typical TRW costs of keypunching as two cents per word for Russian technical text, and one cent per word for English. They also cite cost figures as low as half a cent per word at the CIA-Georgetown Keypunching Center in Frankfurt and at IBM, but this is exclusive of overhead and computer processing (e.g., editing program) costs, so that the one cent figure appears minimal as of today. However, Kochen reports (1963 [327], p. 7): "While keypunching of text cost roughly one cent/word, new means for recording spoken (and written) text using a steno-keyboard tied to a photodisc storing a Stenocode-English dictionary could possibly reduce the cost to 1/3-cent per word."

3/ Swanson, 1962 [584], p. 471.

4/ Bar-Hillel, 1962 [35], p. 418.

5/ O'Connor, 1963 [443], p. 1.

7.2.4 Summary: Potential Advantages as Bases for Evaluation

In view of the difficulties engendered by the underlying core problems, the criticisms that can be brought against tests of "retrieval effectiveness", the general lack of comparative data and standards of measurement, the question of evaluation of automatic indexing procedures largely reduces to the weighing of potential advantages and disadvantages. In the case of such procedures as KWIC and citation indexing, some of these possibilities, both pro and con, have been discussed previously. In general, suggested bases for evaluation reflecting operational considerations may be summarized as follows:

1. **Speed and timeliness**
2. **Relative economy**
3. **Consistency and reliability**[1]
4. **Elimination of the need for further human intellectual effort after initial planning and programming has been done.**
5. **Providing a product that could not otherwise be obtained.**
6. **Ease of updating and revision of indexes so produced.**[2]

From the point of view of possible operational advantages, these may be combined into the single criterion:

> **The achievement of a more effective and more economical balance between the meeting of the objectives of the indexing system and the utilization of available resources.**

---

[1] **Compare McCormick, 1962 [409], p. 182: "A computer is objective in its operations and it can be repetitive. If given a certain amount of information about a document, it is always able to index the document in a consistent manner. This consistency is desired so as to avoid the situations where a person might index a document differently on various occasions, or where it would be indexed differently by another person when there appears to be no good reason for a difference." Note, however, O'Connor's point previously mentioned, (1963 [443], p. 16): "It has been argued that mechanized indexing has the advantage of consistency... However this argument by itself says very little in favor of mechanized indexing. For two humanly produced index sets for a document which differ somewhat may both be quite useful, though imperfect, while the index set which the same program will always reproduce for the same document may be worthless."**

[2] **See, for example, Youden, 1963 [658], p. 332: "The facility with which indexes may be updated and the ease of selecting items for special bibliographies will result in the majority of indexes being computer produced before many years."**

However, the question of the objectives of the system brings us back full circle to the questions of purpose in terms of particular requirements, of quality, and of how to measure either purpose or quality. Thus we may determine that an automatic indexing procedure produces a product at least as rapidly, at least as inexpensively, at least as consistently as human indexing operations would, and with substantially less investment of manpower resources. However, will this product be as useful or as "good" as the human product?

In view of the many caveats about the present quality of indexing systems 1/ and the lack of standards for measuring quality, 2/ it is important to recognize that we should compare the products of automatic indexing methods "not with hand-crafted excellence, but with the average, the routine output of the over-burdened subject analyst working with the deficiencies of any other indexing system". 3/ Such deficiences include the critical question of how well and how consistently the system, whatever it is, is applied in practice by the human analysts.

7.3 Findings with Respect to Inter-Indexer and Intra-Indexer Consistency

Very few objective studies, despite the obvious relationship to the general questions of quality, pertinency, and reliability of indexing, have as yet been made of inter-indexer and intra-indexer consistency. Perhaps the first investigation both to obtain experimental data and to analyze the observed types of failures to achieve correct assignments was that of Lilley. 4/ He took the answers made to 6 questions by 340 students entering a graduate library school, wherein they were asked to write down the subject headings which they would expect to be applied to other books on the same subject as 6 "sample books" in a system such as the Library of Congress card catalog. Lilley reports:

1/ See, for example, in addition to comments by O'Connor and others previously quoted, Helyar, 1961 [262], p. 110: "The general current of feeling of the meeting as reflected both in the papers and in the discussion is that the standard of indexing is not nearly adequate;" Artandi, 1963 [22], p. 1.: "... 'Good indexing' as such has not been defined satisfactorily and is the function of many variables, some known, others not yet identified"; Tritschler, 1963 [610], p. 5: "... 'Good' indexing is extremely difficult to describe and 'perfect' indexing is impossible to define or measure."

2/ See Cleverdon, 1960 [124], p. 429: "The most important requirement in information retrieval is a recognized standard of measurement and after that we need a satisfactory method of measuring. Only when these have been found will it be possible to know for certain whether any new system of indexing or retrieving information is an improvement on previous methods. At present all those trying to solve the problems of information retrieval are working very much in the dark, uncertain as to the real problems and quite unable to apply any measurements to their proposed solutions."

3/ Kennedy, 1962 [311], p. 126.

4/ Lilley, 1954 [360]: See also Vickery, 1960 [626], p. 4.

"A total of 2245 headings were suggested, averaging 1. 1004 headings per book per student. These headings represented 373 different varieties, of which 368 were different from the headings traced on the Library of Congress cards for the sample books... As an average 62. 17 different headings were suggested for each book...

"When the 368 different varieties of incorrect headings were analyzed in accordance with certain criteria that had been set up, it was found that incorrect specificity was a factor in 93. 48%, incorrect terminology in 79. 08% and incorrect form of entry in 72. 28% of the headings... Over half of the incorrect headings (54. 62%) had some combination of two errors, and almost half (49. 73%) could have been converted into 'correct' headings only by changing the level of specificity, and by revising the terminology, and by altering the form...

"It was also found, contrary to the general assumption that failure in specificity almost always means that the reader is approaching his subject from too broad a point of view, that of those headings in which an incorrect level of specificity was a factor... 64. 82% were too broad and 35. 18% were too narrow." 1/

Lilley then asks the rather plaintive question as to what would happen, given that his quite homogeneous group of subjects, all of them college graduates and all seriously interested in librarianship, could come up with more than 62 different headings, on average, for every heading actually used in the catalog, if his test group had included a larger number of subjects with more heterogeneous interests?

In 1961, Macmillan and Welt investigated the duplicate indexing of 171 papers in a limited area of the medical sciences (1961 [389]). In only 18 percent of the cases was the indexing identical or nearly so. About a third of the papers had been indexed so differently that there was no common correlation. For the rest, terms were used in one case that were missed in the other.

Some brief data on inter-indexer consistency is also provided by Kyle (1962 [342]) for two indexers applying her classification system to 246 arbitraily selected French and English items in the field of political science. Of these, 160 were indexed the same way by both indexers, fpr a consistency figure of 70 percent. Tritschler noted that no items were indexed the same way a second time as they were the first, in small-scale experiments involving 20 documents independently indexed by 7 different people. 2/

Painter (1963 [460]), in her study of problems of duplication and consistency of subject indexing of the reports handled by the Office of Technical Services, proceeded by selecting items from the announcement bulletins of agencies contributing to OTS, having these items re-indexed in the various agencies, and comparing the results with the original indexing assignments. At ASTIA, 94 items were re-indexed, with 1, 239 terms having been assigned to them originally and 1, 119 assigned on the re-run. Overall, 62 percent of those terms originally assigned were also assigned the second time, and 69 percent of the second-time terms had also been assigned originally. However, 111 of the starred descriptors (which are of the most significance in the ASTIA system) were used the first time and not the second, while 98 were used the second time but not the first.

1/ Lilley, 1954 [360], pp. 42 and 43.

2/ Tritschler, 1963 [610], p. 5.

At AEC, 96 items were re-indexed to the subject heading scheme used in Nuclear Science Abstracts. There had been 249 headings assigned to these items originally and 406 were assigned on the second run, for an overall consistency rate of 54 percent, but with 53 percent of the headings used the second time not having been used the first. The sample checked at OTS consisted of 32 items to which 346 descriptors had been assigned the first time and 418 the second. The consistency was 65 percent with respect to the first run and 54 percent with respect to the second. Finally, at the National Agriculture Library 99 items were checked, with results showing a high consistency rating and a similarity of indexing between the two runs of 86 percent. Painter concludes:

> "The consistency rates are not encouraging. Apparently there is little difference between preparation for a manual system and that for a machine system. The percentages indicate that there is no significant difference between consistency where two or three headings are assigned and where twelve or sixteen are assigned. Therefore, we are left with the fact that regardless of these variables, consistency rates range between 60 and 72 per cent." 1/

Jacoby and Slamecka report even less encouraging data (1962 [293]). "In general, the inter-indexer reliability was found to be low (in the vicinity of 20 per cent), the intra-indexer reliability somewhat higher (about 50 per cent)." For a series of tests of indexing of a group of chemical patents by three experienced and three inexperienced indexers, they found that the beginners had average matchings among the terms assigned by them to the same documents of only 12.6 percent and that even for the experienced indexers the average percent of matching terms was only 16.3 percent. 2/ In other studies, these investigators have explored the effects of various indexing aids upon the reliability and consistency of indexing, concluding that the use of prescriptive aids such as authority lists improves reliability and inter-indexer consistency from 8 or 9 percent to 33 percent, while those aids such as thesauri and association lists "which enlarge the indexer's semantic freedom of term choice" are detrimental (Slamecka and Jacoby, 1963 [560]).

Rodgers in a study of intra-indexer consistency reports data for the re-indexing, by the same person at a later date, of 60 documents dealing with the United Arab Republic taken from The New York Times. She reports that the average consistency over all 60 documents was 59 percent. 3/ In a further study of inter-indexer consistency, 20 papers from Area 5, ICSI, were key-word indexed by 16 people all of whom were familiar with the subject matter, (although only 8 completed all 20 papers). Results are given in terms of the proportions of the total number of unique words chosen by 100 percent of the subjects (.008) half of them (.14) and only one of them (.52). 4/ Study of the results in terms of the proportion of words selected in common by any pair of these indexers to the total number of different words selected by them both gave a "grand mean agreement for all two-person combinations for the 8 subjects... [of].. 24 percent against all 20 articles." 5/ The mean percentage of overlap between Luhn's word-frequency selection technique (as applied to the same papers) and any one or more indexers who agreed was .15.

1/ Painter, 1963 [460], p. 94.

2/ Jacoby and Slamecka, 1962 [293], p. 16.

3/ Rodgers, 1961 [504], p. 12.

4/ Rodgers, 1961 [503], p. 50.

5/ Greer, 1963 [239], p. 10.

Still further studies of indexer consistency investigated at the Information Systems Operation division of General Electric have just recently been reported (Korotkin and Oliver, 1964 [331, 332]). In particular, the investigators report on the effects of subject matter familiarity and on the use as a job aid of a reference list of suggested descriptors upon inter-indexer consistency. The material for test consisted of 30 abstracts drawn from Psychological Abstracts, to be indexed by 5 psychologists and 5 non-psychologists in two sessions, with and without use of the "job aid". Results in terms of mean percent consistency were reported as follows:

	Session I	Session II
"Group A (Familiar)	39.0%	53.0%
Group B (Non-familiar)	36.4%	54.0%" 1/

Corroborating evidence of a generally low rate of inter-indexer consistency is provided by noting instances of duplicated indexing that may occur in regularly issued announcement bulletins. During current awareness scanning of the DDC (ASTIA) "TAB" in recent months, members of the staff of the Research Information Center and Advisory Service on Information Processing have caught more than 20 cases of duplicate and even triplicate indexing of the same item. (Two examples can be discovered in Figure 8 a and b). For the 52 independent assignments involved, for these items the average inter-indexer consistency is only 46.1 percent.

On the general subject of indexing consistency, Black comments as follows:

> "There have been enough experiments to indicate that there is no consistency, or very little, between one indexing performance by a given individual and another indexing performance, at a later date, by the same individual. The same inconsistency has been discovered among different individuals all indexing the same documents. Thus there is neither inter-indexer consistency nor intra-indexer consistency in any system that depends on human performance." 2/

There can be little doubt that the quality and consistency of most human indexing, practically available today, is not good. Much of it, because of time and other pressures, is either directly a word-extraction process, or it is inconsistent in assignment of many relevant descriptors and subject category labels. On the other hand, today's indexing, whether accomplished by man or machine, is probably no better and no worse than any other classificatory or indexing procedures. The only excuse, therefore, for choice between man and machine is the cost/benefit ratio which is related on the one hand to specific operational considerations and on the other to the question of whether or not various indexers, and various users, would agree with the machine as much as they agree with each other.

Before turning to some of the operational considerations affecting the cost-benefit ratio, however, certain special factors should be briefly mentioned.

7.4 Special Factors and Other Suggested Bases for Evaluation

The difficulties and problems of evaluation so far considered are generally applicable to any indexing system, whether manual or automatic. Certain special factors arise, however when we consider some of the proposed automatic assignment and automatic classification techniques. In addition, the prospects for computer processing hold at least the

---

1/ Korotkin and Oliver, 1964 [331], p. 7.

2/ Black, 1963 [64], pp. 16-17.

AD-408 841 Div. 32

Joint Publications Research Service, Washington, D. C.
UNDERWATER FISHERY RESEARCH IN THE USSR, by V. P. Zaitsev. 2 Apr 63, 14p. 18501
Unclassified report

Trans. of Okeanologiya (USSR), 1962, v. 2, no. 6, pp. 961-969. Also from OTS for $.50 as rept. 63 21481.

Descriptors: (*Fishes), Scientific research, (*Oceanology), Marine biology, Ocean currents, Diving, (*Oceanographic equipment), (*Underwater equipment), Submarines.

AD-408 849 Div. 32

Joint Publications Research Service, Washington, D. C.
TOWARD NEW PROGRESS OF SCIENCE AND TECHNOLOGY AND IMPORTANT PROBLEMS OF SCIENTIFIC ORGANIZATION, by M. V. Keldysh and M. A. Lavrent'ev. 20 May 63, 25p. 19283
Unclassified report

Trans. of Akademiya Nauk SSSR. Vestnik, 1962, v. 32, no. 12, p. 9-14 and 16-18. Also from OTS for $.75 as rept. 63-21864.

Descriptors: (*Scientific research). (*Scientific organizations). Energy management, Materials, Semiconductors, Chemical industry, Agriculture, Computers.

AD-408 854 Div. 32

Joint Publications Research Service, Washington, D. C.
ORDER CONCERNING COMMISSION FOR USE OF UNIVERSE FOR PEACEFUL PURPOSES NO. 36.
29 Apr 63, 2p. 18954.
Unclassified report

Trans. from Sluzbeni List, Belgrade (Yugoslavia) 1963 19:12, p. 163. Notice: Also from OTS for $.50 as rept. 63 21705.

Descriptors: (*Space flight), (*Political science), Scientific organizations.

AD-408 866 Div. 32

Joint Publications Research Service, Washington, D. C.
THE PAST TEN YEARS AT VINITI (ALL-UNION INSTITUTE OF SCIENTIFIC AND TECHNICAL INFORMATION), by V. A. Polushkin. 29 May 63, 3p. 19482
Unclassified report

Trans. of Akademiya Nauk SSSR. Vestnik, 1963, v. 33, no. 3, pp. 127-128. Also from OTS for $.50 as rept. 63 21950.

Descriptors: (*Scientific organizations), Documentation, (*Communication theory).

AD-408 877 Div. 32

Joint Publications Research Service, Washington, D. C.
ABSTRACTS FROM EAST EUROPEAN SCIENTIFIC AND TECHNICAL JOURNALS NO. 190 (BIOLOGY AND MEDICINE SERIES).
29 May 63, 27p. 19470
Unclassified report

Consists of abstracts of articles from selected scientific and technical journals of Bulgaria, Poland and Yugoslavia. Also from OTS for $.75 as rept. 63-21948.

Descriptors: (*Abstracts), Bibliographies, (*Biology), (*Medicine), Genetics, Blood, Drugs, Pharmacology, Microorganisms, Biochemistry, Diseases, Neurology, Therapy, Medical examination, Vaccines, Viruses, Plants (Botany), Scientific personnel, Toxicity.

AD-408 878 Div. 32

Joint Publications Research Service, Washington, D. C.
ABSTRACTS FROM EAST EUROPEAN SCIENTIFIC AND TECHNICAL JOURNALS NO. 187 (BIOLOGY AND MEDICINE SERIES).
29 May 63, 20p. 19465
Unclassified report

Consists of abstracts of articles from selected scientific and technical journals of Hungary. Also from OTS for $.75 as rept. 63-21945.

Descriptors: (*Abstracts), Bibliographies, (*Biology), (*Medicine), Chemical analysis, Drugs, Neurology, Surgery, Wounds and injuries, Pathology, Diet, Public health, Infants, Toxicity.

AD-408 887 Div. 32

Joint Publications Research Service, Washington, D. C.
CYBERNETIC MACHINES: SELECTED ARTICLES.
27 Aug 62, 15p. 14962
Unclassified report

Trans. from Leninskoe Znamya (USSR) 1962, July; Literaturnaya Gazeta (USSR) 1962, 7 July; Pravda, Moscow (USSR) 1962, 5 July. Also from OTS for $.50 as rept. 62-11760.

Descriptors: (*Cybernetics), (*Digital computers), Learning, Computer logic, Design.

Contents:
''Thinking'' machines: friends or enemies, by V. Trapeznikov
Machine rns to learn, by G. Zelenko
Can a machine create a design, by Yu. Sinyakov

AD-408 937 Div. 32
(TISTB/PCR)

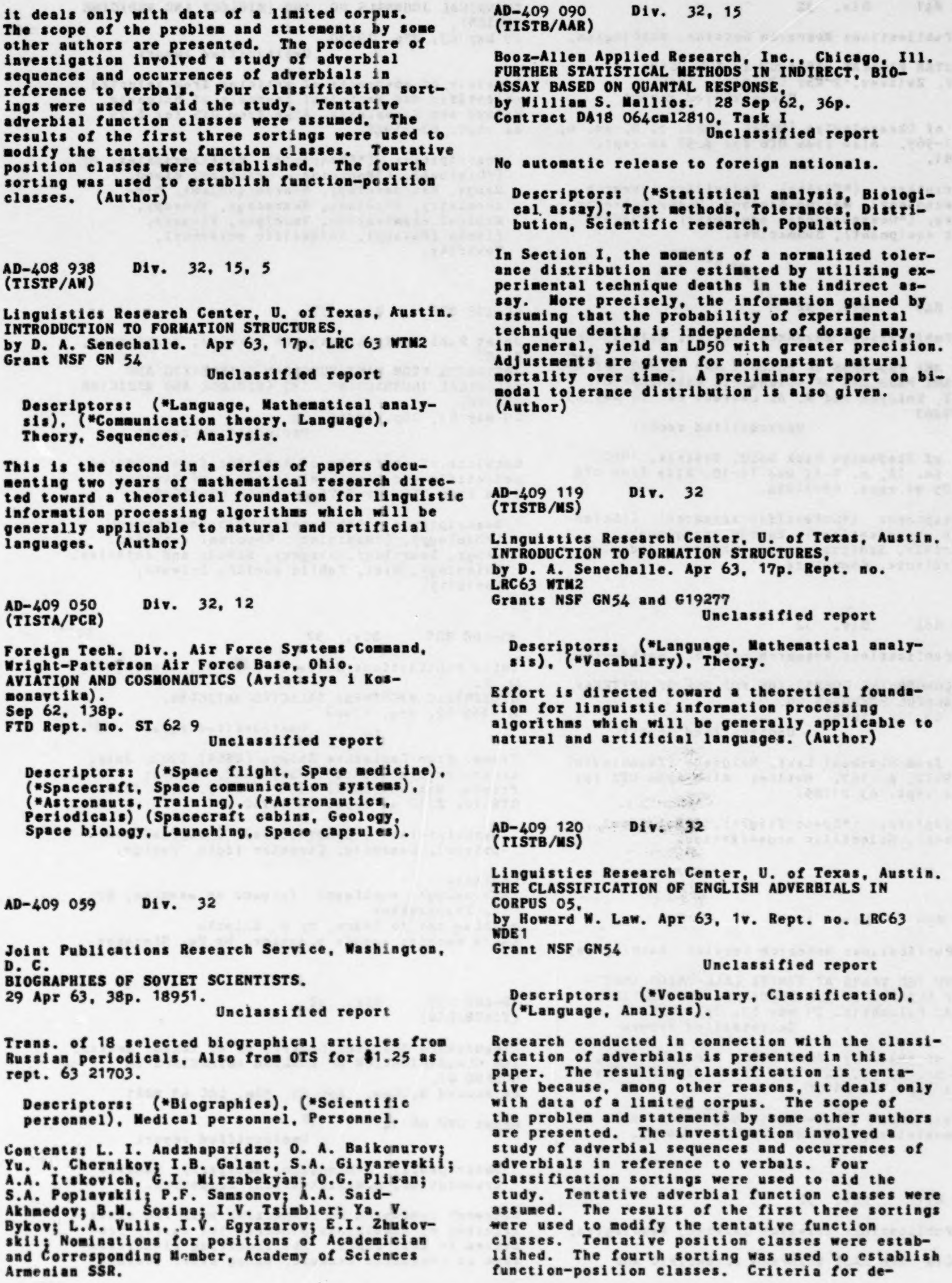
Linguistics Research Center, U. of Texas, Austin.
THE CLASSIFICATION OF ENGLISH ADVERBIALS IN CORPUS 05,
by Howard W. Law. Apr 63, 52p. LRC 63 WDE1

Grant NSF GN 54
Unclassified report

Descriptors: (*Language, Analysis), Machine translation, Classification, Computers.

Research conducted in connection with the classification of adverbials produced the survey presented in this paper. The resulting classification is tentative because, among other reasons,

Figure 8a. Examples of Duplicate Indexing



it deals only with data of a limited corpus. The scope of the problem and statements by some other authors are presented. The procedure of investigation involved a study of adverbial sequences and occurrences of adverbials in reference to verbals. Four classification sortings were used to aid the study. Tentative adverbial function classes were assumed. The results of the first three sortings were used to modify the tentative function classes. Tentative position classes were established. The fourth sorting was used to establish function-position classes. (Author)

AD-408 938 Div. 32, 15, 5
(TISTP/AW)

Linguistics Research Center, U. of Texas, Austin.
INTRODUCTION TO FORMATION STRUCTURES,
by D. A. Senechalle. Apr 63, 17p. LRC 63 WTM2
Grant NSF GN 54
Unclassified report

Descriptors: (*Language, Mathematical analysis), (*Communication theory, Language), Theory, Sequences, Analysis.

This is the second in a series of papers documenting two years of mathematical research directed toward a theoretical foundation for linguistic information processing algorithms which will be generally applicable to natural and artificial languages. (Author)

AD-409 050 Div. 32, 12
(TISTA/PCR)

Foreign Tech. Div., Air Force Systems Command, Wright-Patterson Air Force Base, Ohio.
AVIATION AND COSMONAUTICS (Aviatsiya i Kosmonavtika).
Sep 62, 138p.
FTD Rept. no. ST 62 9
Unclassified report

Descriptors: (*Space flight, Space medicine), (*Spacecraft, Space communication systems), (*Astronauts, Training), (*Astronautics, Periodicals) (Spacecraft cabins, Geology, Space biology, Launching, Space capsules).

AD-409 059 Div. 32

Joint Publications Research Service, Washington, D. C.
BIOGRAPHIES OF SOVIET SCIENTISTS.
29 Apr 63, 38p. 18951.
Unclassified report

Trans. of 18 selected biographical articles from Russian periodicals, Also from OTS for $1.25 as rept. 63 21703.

Descriptors: (*Biographies), (*Scientific personnel), Medical personnel, Personnel.

Contents: L. I. Andzhaparidze; O. A. Baikonurov; Yu. A. Chernikov; I.B. Galant, S.A. Gilyarevskii; A.A. Itskovich, G.I. Mirzabekyan; O.G. Plisan; S.A. Poplavskii; P.F. Samsonov; A.A. Said-Akhmedov; B.M. Sosina; I.V. Tsimbler; Ya. V. Bykov; L.A. Vulis, I.V. Egyazarov; E.I. Zhukovskii; Nominations for positions of Academician and Corresponding Member, Academy of Sciences Armenian SSR.

AD-409 090 Div. 32, 15
(TISTB/AAR)

Booz-Allen Applied Research, Inc., Chicago, Ill.
FURTHER STATISTICAL METHODS IN INDIRECT, BIOASSAY BASED ON QUANTAL RESPONSE,
by William S. Mallios. 28 Sep 62, 36p.
Contract DA18 064cml2810, Task I
Unclassified report

No automatic release to foreign nationals.

Descriptors: (*Statistical analysis, Biological assay), Test methods, Tolerances, Distribution, Scientific research, Population.

In Section I, the moments of a normalized tolerance distribution are estimated by utilizing experimental technique deaths in the indirect assay. More precisely, the information gained by assuming that the probability of experimental technique deaths is independent of dosage may, in general, yield an LD50 with greater precision. Adjustments are given for nonconstant natural mortality over time. A preliminary report on bimodal tolerance distribution is also given. (Author)

AD-409 119 Div. 32
(TISTB/MS)

Linguistics Research Center, U. of Texas, Austin.
INTRODUCTION TO FORMATION STRUCTURES,
by D. A. Senechalle. Apr 63, 17p. Rept. no. LRC63 WTM2
Grants NSF GN54 and G19277
Unclassified report

Descriptors: (*Language, Mathematical analysis), (*Vacabulary), Theory.

Effort is directed toward a theoretical foundation for linguistic information processing algorithms which will be generally applicable to natural and artificial languages. (Author)

AD-409 120 Div. 32
(TISTB/MS)

Linguistics Research Center, U. of Texas, Austin.
THE CLASSIFICATION OF ENGLISH ADVERBIALS IN CORPUS 05,
by Howard W. Law. Apr 63, 1v. Rept. no. LRC63 WDE1
Grant NSF GN54
Unclassified report

Descriptors: (*Vocabulary, Classification), (*Language, Analysis).

Research conducted in connection with the classification of adverbials is presented in this paper. The resulting classification is tentative because, among other reasons, it deals only with data of a limited corpus. The scope of the problem and statements by some other authors are presented. The investigation involved a study of adverbial sequences and occurrences of adverbials in reference to verbals. Four classification sortings were used to aid the study. Tentative adverbial function classes were assumed. The results of the first three sortings were used to modify the tentative function classes. Tentative position classes were established. The fourth sorting was used to establish function-position classes. Criteria for de-

Figure 8b. Examples of Duplicate Indexing

promise of more objective measures of performance or quality than evaluative techniques available today.

Examples of the special factors involved in assignment indexing techniques and automatic classification include the question of the amount of computation required in the inversion and other manipulations of large matrices 1/ and the concommitant problems of how large a vocabulary of clue words can be used effectively and of whether some documents cannot be indexed at all because they contain none of these words. 2/ There is, as Needham says, "no merit in a classification program which can only be applied to a couple of hundred objects." 3/

In the various techniques for automatic clustering or categorization of documents, there are serious questions of whether the groupings can be conveniently named or displayed for the benefit of the user. 4/ Another example of special factors in the appraisal of an automatically generated classification scheme is as follows:

> "Operational testing is displeasing in that it puts off any verification until right at the end; it is expensive; there is not much experience on how to do it in a realistic way; and it is ill-controlled in the sense that the practical performance of a system is influenced by many other factors than the classification it embodies." 5/

Examples of suggested bases for evaluation made possible by machine processing itself include proposals by Doyle and Garvin, among others. Doyle in particular suggests the substitution for the elusive concept of "relevance" of criteria based on "sharpness of separation of exploratory regions in which the searcher finds documents of interest from those in which he does not find such documents." 6/ He further emphasizes the need for discriminating a particular document from other topically close documents (Doyle, 1961 [166]) and suggests that "this decision can never be made by a human---only by a computer, which is the only agency capable of having full consciousness of the contents of a library." 7/ Garvin considers the more general problems of language and meaning, and suggests that there are two kinds of "observable and operationally tractable manifestations of linguistic meaning", ---namely, translation and paraphrase, and that these may be investigated by techniques of linguistic data processing. 8/ Edmundson, however, points out that while there is in general only one translation of a document, there may be as many abstracts (and, by implication, index sets) as there are users. 9/ Thus we are back again at the questions of purpose and relevance.

1/ Compare Williams, 1963 [642], p. 162.

2/ See Maron and Borko, various references.

3/ Needham, 1963 [433], p. 8.

4/ See, for example, Doyle, 1963 [162], p. 6: "Several researchers have tried to group topically close articles, usually by statistical means, but it is rather difficult to get any benefit from this grouping unless you can represent these groups for human inspection."

5/ Needham, 1963 [432], p. 2.

6/ Doyle, 1963 [164], p. 200.

7/ Doyle, 1961 [169], p. 23.

8/ Garvin, 1961 [224], p. 137.

9/ Edmundson, 1962 [178], p. 4.

8. OPERATIONAL CONSIDERATIONS

Whatever the verdict of evaluation of one or more automatic indexing techniques, whether of the derivative, modified derivative, or assignment type, there are certain operational considerations and problems that typically affect any attempt to apply such techniques in actual production operations. These considerations, which also affect linguistic data processing operations in general, include input considerations, availability of methods or devices for converting text to machine-usable form, programming considerations, questions of format and content of output, and problems of customer acceptance of the machine products.

8.1 Questions of input

Input considerations include, first, questions of the extent and availability of material which can be handled directly by the machine. This may be limited to title only, to title plus abstract, title plus other material, 1/ preselected text or automatically generated extracts; or it may in a few cases extend to full running text. Possible future requirements may extend to the processing not only of full text but of interspersed graphic material (equations, charts, diagrams, drawings, photographs) as well.

We have considered typical arguments for and against the limitation of input to titles only, to augmented titles, and to abstracts in other sections of this report. The points to be emphasized here are requirements for pre-editing or post-editing, provisions for error detection and error correction, the time and cost requirements of conversion equipment if material is not already available in machine-usable form, and the like. As Cornelius suggests:

> "Present day computers, if used for machine indexing, will be generally input limited and will require excessive data preparation. Causes of these limitations are: time required for translation to machine language, verification of this machine language, and the capability or lack of capability of correction in the input media." 2/

Examples of pre-editing requirements, even for the simple case of keyword-in-title indexing, include the spelling out of chemical symbols, the encoding or the omission of subscripts and superscripts, insertions of hyphens to prevent indexing of a word, and substitutions of blanks for hyphens in compound words to assure indexing of each component. 3/ For full text, a far more extensive and elaborate set of rules and conventions must be developed and applied. 4/ Other editing may be required for format standard-

---

1/ This may specifically include cited titles, as suggested variously by Bohnert, 1962 [69], p. 19; Giuliano and Jones, 1962 [229], p. 10; Swanson, 1963 [580], p. 1; Gallagher and Toomey, 1963 [205], p. 53; and as used in the SADSACT method, see pp. 98-99 of this report.

2/ Cornelius, 1962 [140], p. 42.

3/ See, for example, Kennedy, 1961 [311], p. 120.

4/ See, for example the sophisticated proposals of Nugent, 1959 [441], and Newman et al, 1960 [439].

ization, especially in the case of citation indexes compiled by machine. 1/ O'Connor notes, however, that "the provision of pre-editing information can slow down the keypuncher or typist, increase the chance of mistakes, and require more intelligence or training on the typist's part." 2/

Questions of error detection and error correction apply both to the original text and to transcribed versions if these are necessary. That is, the basic documents themselves may contain typographical errors, misspellings, and the like, and additional errors are bound to occur at all subsequent stages requiring human processing. Wyllys discusses the need for the correction of spelling errors, mentions suggested computer programs for detection, and cites a private communication from Stiles suggesting that the criteria for accepting words as valid be either that they are identified as already being in the system vocabulary or that they occur at least twice in the input item. 3/

Swanson's analysis of the reasons for retrieving irrelevant, and failing to retrieve relevant, material in the case of text searching on the nuclear physics abstracts includes typical data on the effect of errors. 4/ He found, for example, that failures to record hyphenated words, subscripts, superscripts and other special symbols accounted for about 5 percent of failures to retrieve relevant items, and errors in transcription of either text or search instructions accounted for another 3 percent of these failures. Errors in keypunching of the search requests alone accounted for 4 percent of the cases of irrelevant retrievals. By contrast, in the newspaper clippings experiments where the input material was already in machine-usable form transcription errors were not a factor but the input tape itself had many errors. In this special case, however, Swanson reports: "Garbles are not important simply because messages are sufficiently redundant to insure that even if one or two keywords for a given category are garbled, almost invariably others are present." 5/

The news clippings material used by Swanson represents one class of materials that are today initially available in machine-usable form, because the original recording of the message or text resulted in a machine-usable medium, such as punched paper tape. A punched paper tape is produced as the product of many typesetting operations, especially for newspaper and magazine publication, and this will be increasingly true in the future, together with computer-prepared tapes for input to automatic typographic composing equipment. To date, however, equipment to convert from these tapes to the particular machine language of a given computer processing system is largely non-available, is costly, and is highly subject to error. 6/

1/ See, for example, Atherton, 1962 [25], p. 4; Marthaler, 1963 [399], p. 22. However, at least one computer program has been developed to assist in this process. See Thompson, 1963 [600], p. II-1: "The present program takes bibliographic citations and automatically arranges then into a standard format in such a way that the various parts of the citation are unambiguously identified. These standardized citations can later be processed by sorting and matching procedures to identify similar citations and to effect various rearrangements."

2/ O'Connor, 1960 [444], p. 8.

3/ Wyllys, 1963 [653], p. 15.

4/ Swanson, 1961 [586], Appendix.

5/ Swanson, 1963 [580], p. 5.

6/ Compare, for example, Savage, 1958 [521], p. 11: "The use of tape as the original input to the process has offered a number of problems which have yet to be solved. One is the occurrence of typographical errors."

Moreover, to date, very little material in the scientific and technical literature is available in this form. As of 1961, it was reported that a survey by McGraw-Hill indicated that only about 2 or 3 percent of the publications in the United States were then prepared by typesetting tape, that most of this was in the form of Monotype tape which because of its 30-column width and special format is not generally compatible with tape reading equipment, and that tapes had many errors in them which would require considerable effort to correct. 1/ As of late 1963, Bennett reports:

> "Computer processing of natural language text material requires that a body of data be available in machine-readable form. At present such a body of data results only from a direct human copying process. An inquiry into existing transcriptions of text which were machine-readable showed that they were abbreviated both in terms of completeness and in number of symbols represented. As an alternative text produced as a by-product of typesetting operations is clearly an eventual possibility, but present practices make the detection of unit delimiters such as ends-of-sentences difficult." 2/

In the future, both machine-usable text from publishers and printers and the similarly machine-usable paper tape produced as a byproduct from the original keystroking of manuscript on such equipment as Flexowriters and Justowriters may alleviate this problem for new items. Nevertheless, the wealth of the world's present literature, the informal and unpublished technical reports of high current interest but limited initial distribution, and material acquired from foreign sources, will continue to pose for the foreseeable future major problems either of automatic reading of the printed page or of human retranscription at high cost.

While there have been many promising developments in automatic character recognition techniques, the devices that are now available for production use are limited to small character sets, such as a single alphabet in a single font, often of special design. The multi-font page reader is not only not yet commercially available but may not become so for some years to come. Even if it were, there are many unresolved and as yet incompletely specified problems involved in the development of suitable rules for the machine so that it can distinguish between title or page number and text, figure caption and text, author's name in a cited reference and the title of the paper cited, and the like. A case in point, not only for automatic reading equipment of the future but for machine processing of machine-usable material available today, is the difficulty of machine recognition of punctuation marks as used for different purposes. 3/

In the absence, then, both of scientific and technical documents already in machine language form and of character recognition equipment capable of reading the printed page, we are left with the unsatisfactory situation of re-transcribing input material either by use of a tape typewriter or by keypunching to punched cards. That this situation is unsatisfactory and is a major bottleneck in machine processing of text in excess of the bibliographic citation data only is evidenced by such typical statements as these:

1/ Cornelius, 1962 [140], p. 47.

2/ Bennett, 1963 [50], p. 141.

3/ See Bennett quotation above; Luhn, 1959 [384], p. 22, and Coyaud, 1963 [143].

> "The expense of transcribing such documents in their entirety will be justifiable to a limited extent only and it may, therefore, be assumed that automatic processing will be mainly applied to future literature." 1/
>
> "As long as we are limited to using the equipment that is available now, the preparation of data for input will be an expensive procedure and a major cost factor in automatic processing of natural language." 2/
>
> "... In a discussion of indexing by machine, we must recognize the preparation of input to the system as the major item of cost of operation." 3/
>
> "Present inability to read documents automatically would make it necessary to punch cards or tapes, an operation likely to be even more expensive than reading by humans." 4/

In addition to the high costs of manual retranscription, it is also noted that keypunching "tends to undermine the purpose of natural text retrieval by requiring human effort at the input end of the process." 5/

In particular, keypunching or keystroking requirements undermine the purposes of rapid indexing as well as filing for retrieval by virtue of the time required to transcribe text. Horty and Walsh report, for example:

> "Flexowriter operators can produce between 1400 and 1800 lines per day of statutory text. Keypunch operators used in previous experiments could punch approximately 100 lines per hour of alphabetic materials, but could not maintain this rate for a sustained period of time." 6/

Thus, until such time as more versatile character recognition equipment is available, even some of the most ardent advocates of full text processing are forced to the use of considerably less than full text for other than research purposes. Swanson comments, for example:

> "... One must note that the manual recording of text may be exorbitantly expensive. If so, a judicious selection process may permit a reasonable compromise between the expense of input and the depth of indexing which results. For example, it is reasonable to select the title, abstract, table of contents (if any), sub-headings, and key sentences or paragraphs." 7/

1/ Luhn, 1959 [384], p. 2.

2/ Ray, 1961 [496], p. 51.

3/ Howerton, 1961 [282], p. 327.

4/ Levery, 1963 [359], p. 235.

5/ Doyle, 1959 [168], p. 2.

6/ Horty and Walsh, 1963 [280], p. 259.

7/ Swanson, 1963 [580], p. 1.

"Costs come much more into line if we make available to the machine something on the order of one per cent of the full text. Then, of course, the problem of selecting that one per cent presents itself." [1]

8.2 Examples of Processing Considerations

A second major area of operational considerations involves the machine processing problems, given a specified input. For most of the automatic derivative, and modified or normalized derivative, schemes, this is primarily a question of the limitations of machine language to a vocabulary of, typically, no more than 64 distinct characters for input, internal manipulation, and output. In addition, the limited number of characters that can be packed into a single machine-word complicates internal processing, storage, file look-up (i.e., against exclusion or inclusion lists), and sorting operations.

Arbitrary truncation of text words to, say, 6 characters per word, leads to certain computer processing or storage economics. However, it leads also to complications in the selection of words either to be included (clue word lists) or excluded (stop lists) in many of the proposed methods both for derivative and for assignment indexing. Additional problems of artificial homography are created. Obvious examples are "Probab-le, -ility"; "Condit-ion, -ional, " "Freque-nt, -ntly, -ncy, " "Commun-ity, -ication;-al", and the like. Barnes and Resnick include in their studies of the effectiveness of an SDI System [2] the use of 6 different truncation levels (from 4 to 9 characters). No significant differences were found in terms of the number of hits (matches of a new item to a user's profile which he considered to be of definite interest to him) but there were significant differences in the number of notifications sent him, as presumably matching his interest, and the amount of "trash" (irrelevant items) among these notifications.

The importance of the selection criteria in derivative indexing, operationally considered, is largely a matter of the length and the contents of the stop lists. Variability in practice among the various producers of KWIC indexes has previously been noted, [3] but there are some interrelated and interlocking factors which affect the quality, the costs, and the customer acceptance of this type of machine-generated index. First, the number of pages in a printed index is directly related to the total costs of producing that index. [4] The amount of material covered on a single page can be increased by photographic or other type of reduction (e.g., the 96 lines per page of the Bell Laboratories KWIC program output are reduced by xerography to 62 percent of the machine output page size), (Kennedy, 1961 [311]) but the reduction must not be such as to exceed reasonable limits of legibility.

This, in turn, means that the number of entries generated for each title (obviously, a function of the words that survive stop list purging) needs to be held to a reasonable minimum. Thus:

> "One of the major limitations of the published index stems from the conflict between the quantity of text that must be placed between the covers and the capacity of the printed page to handle it. The size of the page and the legibility of the printing determines the maximum density of characters which can be read without special aids." [5]

---

[1] Swanson, 1962 [584], pp. 470-471.

[2] Barnes and Resnick, 1963 [36]. See also p. 148 of this report.

[3] See discussion, pp. 65-66.

[4] See Markus, 1963 [394], p. 16.

[5] Taine, 1961 [592], p. 153.

The question of stop list effectiveness therefore becomes an operational factor as well as one that may affect the quality and acceptability of the product. On the other hand, too generous a purging of the input titles may of course reduce the utility of the title index by the elimination of too many potential access points and, in particular, many that users may be most tempted to look for.

A related problem has to do with the number of pages required because of the length of the title line allowed in the listings. A suggestion advanced by Brandenberg (1963 [80]) is the assignment of numeric codes to the machine stop words used and the insertion of these codes into the listed title line in the place of these presumably insignificant words. Thus one of the KWIC entries for the title, "Determining Aspects of the Russian Verb from Context in Machine Translation" might go from:

```
RMINING ASPECT OF THE   CONTEXT IN   MACHINE TRANSLATION.  /DETE to:
ERMINING 032 416 712 RUS   CONTEXT 308   MACHINE TRANSLATION.  /DET
```

This particular example was picked at random from a KWIC index utilizing a 103-106 character title line, 1/ but it was deliberately shortened to the 60-character line length found in many such indexes in order to illustrate effects of chopping and wrap-around. Coincidentally, it also illustrates some of the difficulties of designing a well-balanced exclusion list since in this case the purged word "aspect" is apparently being used in a technical sense rather than in the common one of "Various aspects of...". By accident, this case does show rather severe "aspects" of the chopping problem in the loss also, for this entry, of "Russian" and "verb" although they would of course be picked up in the entry blocks for these words. Certainly, however, the claimed advantages of context checking are not striking, even without the introduction of the numeric codes. It is true that for excluded words longer in length than those in our example the possible conservation of the character-space to reduce the chopping effects for the same length line may result in improvements. However, the replacement of, for example, "Preliminary investigations of..." by numeric codes would hardly assist the user in determining quickly from the many possible entries under "..." which he should select for further personal perusal.

Turning to the case of automatic assignment indexing, the processing considerations likely to be involved in operational factors affecting the evaluation of a system are much less easily exemplified. Obviously, conditions that hold for research experiments on small (and usually, especially selected) samples do not necessarily relate to requirements in potential productive applications. Exceptions are the problems of the sizes of term-term and term-document co-occurrence correlation matrices that can be readily manipulated, previously mentioned, 2/ and the concurrent problems of the size, and hence the representativeness, of inclusion lists or clue-word vocabularies that can be accommodated.

Both Maron and Borko found, even in their limited test samples, a certain proportion of new items that could not be indexed or categorized at all because these new items did not contain any of the clue words recognizable by the system. 3/ Due perhaps to longer selective clue word lists, as well as to the special nature of his items, Swanson found no instances, for 775 test items, of failure to assign because of lack of indicative clues in the input material. In the case of 60 tests against the SADSACT model, which uses approximately 1, 600 words drawn from a "teaching sample" of items previously indexed to descriptors, (related by frequency of co-occurrence to any of 70-odd descriptors with whose

---

1/ Walkowicz, 1963 [629], pp. 136 and 137.

2/ See pp. 108 and 160 of this report.

3/ See Maron, 1961 [395]; also Borko and Bernick, 1963 [78].

assignment they had co-occurred), the machine had a sufficient basis in the input material for the derivation of a selection-score for at least 12 descriptors for each new item. The items were closely similar to, though not identical with, the source items from which the word associations with descriptors assigned had been drawn. The sample is obviously critically small. Nevertheless, the possibility that extensive clue word lists, notwithstanding the incorporation of trivial and even erroneous associations, can be used as effectively as smaller, more precise, and more carefully tailored lists, but with significant gains in memory space or computational requirements, is suggestive. A somewhat related conclusion, again reflecting the effect of processing requirements, is stated by Needham as follows:

> "The main point to be made is that theoretical elegance must be sacrificed to computational possibility: there is no merit in a classification program which can only be applied to a couple of hundred objects." 1/

In KWIC type derivative indexing by machine, except in terms of allowable character sets and word-lengths conveniently processed, the problem of appropriate programming languages does not arise to any serious extent. For the processing of material in research on natural language text, however, the choice of interpretative and compiler types of automatic programming languages may involve computational requirements which, while being inappropriate in a production situation, offer considerable flexibility and versatility for experimental purposes. Examples of special programs of this type include the use of Yngve's COMIT by Baxendale and Knowlton, the development and use of FEAT by Olney, Doyle, and others at SDC, and the use of list-processing techniques in the General Inquirer system. 2/ Yngve describes the use of his program as follows:

> "COMIT has also been used in the experimental work in information retrieval of Baxendale and Knowlton at IBM. The purpose of their COMIT program was to accept as input the title of a document and to produce as output, not only descriptors, but pairs of descriptors which are roughly of the form adjective-noun. The purpose of the work is to automatically generate, from document titles, retrieval words of a more specific nature than simply Boolean functions of the existence of certain words in a title." 3/

The FEAT program was designed originally for word and significant-word-pair frequency counts. Olney describes the program in part, as follows:

> "FEAT is designed to perform frequency and summary counts of words and word pairs occurring in its natural text input; i.e., text written in ordinary English and transcribed into Hollerith code according to some set of keypunching rules. To focus attention on the semantic aspects of word pairs rather than on their syntactic aspect, pairs of which one member is a function word, such as 'the', 'is', 'by', etc., are excluded."
>
> "Using a bucket list structure of the type proposed by C. J. Sheen in FN-1634, the program sorts each incoming word serially, constructing a list within each of 256 buckets for good words of a given alphabetic range ... and another list within each good word entry for the Doubles and Reverses which will be ordered alphabetically

---

1/ Needham, 1963 [433], p. 8.

2/ Stone, et al, various references, p. 137 of this report.

3/ Yngve, 1962 [655], p. 26.

> on that word ... If there are four different Double types of which the first word is 'external' the addresses of the four different second words form a new list which is linked to the entry for 'external'. Each word type occurs only once in core, and all word pairs of which it is a member refer to it by means of its core addresses."
>
> "The program could process millions of words, automatically generating frequency counts far larger than the Thorndike and Lange counts, which cost many man-years, and in addition, FEAT would provide complete lists of word pairs (Doubles and Reverses), which, so far as we know, have never been counted in a sample of appreciable size, despite their importance for semantic analysis of text."

FEAT is used, together with a modified version of the Proto-Synthex program, and special output formatting routines, for another SDC program, the Descriptor Word Index Program, which produces a content-word-concordance for natural language text as well as statistics reflecting the type of words that occur, frequencies of occurrence, and positional data, (Olney, 1960 [457], 1961 [456]; Stone, 1962 [574].

The IPL-V list-processing language is used by Kochen in some of his work on simulated concept processing by machine. Programs for accepting sentences written in a formal language which was constructed of names and logical predicates (inserted either from a console or in the form of punched cards), for updating and re-organizing a file of such sentences, for storing and manipulating metalinguistic sentences such as "If X is author of Y and Y pertains to topic Z, then X has worked on Topic Z", for interrogating the file, and for tracing associations between names linked through various predicates, have been written in this language. 1/

8.3 Output Considerations

Turning to operational problems of output, the question of limitations of computer printout language to, in most cases, a single set of upper case alphabetic characters, numerals, and a few special symbols, 2/ is a serious factor in customer acceptance with respect to appearance -- format, legibility, readability. Involved here are questions previously mentioned. Where, in the only presently available outputs of machine-generated indexes, the KWIC type permuted title indexes, should the indexing access point "slot" be on the page? Should all or only part of the title be displayed? Should 60- or 106-character lines be used? More detailed discussion of these and related points are provided by, for example, Youden (1963 [658]) Kennedy (1962 [311]) and Brandenberg (1963 [80]).

A separate, but related question, is how much identification, and in what form, should be provided for the item itself either directly as a part of the index entry or by cross-reference to the address of more detailed information. There seems to be quite general agreement that the typical user needs something more than author's name and title

---

1/ Kochen, et al, 1962 [328], p. 34.

2/ See, for example, Lipetz, 1960 [365], p. 252: "A disadvantage of keypunched cards, however, is the lack of capacity to record or to print other symbols than a one-case alphabet, one case of arabic numerals, and about a dozen punctuation marks and miscellaneous symbols. Citations in the scientific literature generally make use of a much larger number of significant symbols: multiple cases, multiple fonts, italics, boldface, Greek letters, mathematical symbols, etc." Note, however, that Chemical-Biological Activities, a digest produced by Chemical Abstracts Service, uses printouts of the modified IBM 1403 chain printer, using 120 characters (see Fig. 5).

alone to guide him. 1/ However, if the full bibliographic citation, perhaps the abstract as well, is to be printed out by machine, the problems of limited character set are even more severe. This problem is today being solved, in some cases, by separate operations involving sorting and assembly of the full citations and abstracts of the items indexed, separately prepared, for photographic reproduction or typesetting. Hopefully, this partial solution will become obsolete as automatic type-composition equipment and computer-prepared typesetting techniques become more generally available.

Operational considerations thus involve the costs, the availability, and the limitations of equipment now usable for machine-generated index production. Schultz and Schwartz report, as of October, 1962.

> "There are two major bottlenecks in automated index production caused by inadequate equipment development at the present state-of-the-art:
>
> "1. There is no way of using automatic input of the printed page or the indexer's notes;
>
> "2. There is insufficient flexibility in the forms of output available for a computer-produced index.
>
> Both of these areas are being worked on by equipment manufacturers, and an early solution has been promised." 2/

In general, operational considerations of this type do not affect the appraisal of automatic assignment indexing techniques, because these have not yet been developed to the point of practical application on any realistic scale. Moreover, the difficulties of problem definition and basic understanding of language and meaning yet remaining to be resolved are such that radical new advances in computer technology, associative memories, character readers and pattern recognition devices may completely alter the picture before practical systems are ready for operational tests. Thus, for example, it is claimed:

> "It appears desirable to begin experimentation with automatic indexing so that solutions will become known by the time character recognition equipment will have passed the laboratory stage." 3/

Similarly, Doyle suggests that the "present rate of solution of the intellectual problems of IR is sufficiently slow that these advanced devices will be in common use long before IR will truly benefit from their presence", and he urges that researchers proceed as though such machines were already with us. 4/

1/ Compare, for example, Montgomery and Swanson, 1962 [421], p. 366: "This study suggests that indexing should be based on more than titles and that a bibliographic citation system should present to the requestor something more than titles"; See also, in addition to references cited, p. 61, footnote 1, IBM "ACSI-matic auto-abstracting project...", Vol 3, 1961 [290], p. 89: "The use of titles in document searching without any additional abstract seems to lead to a high number of ... errors, i.e., accepting documents which should be rejected, as not enough information is available to judge the pertinence of documents."

2/ Schultz and Schwartz, 1962 [531], p. 432.

3/ Levery, 1963 [359], p. 235.

4/ Doyle, 1961 [169], p. 3.

9. CONCLUSION: APPRAISAL OF THE STATE OF THE ART IN AUTOMATIC INDEXING

Notwithstanding the difficulties of evaluation we have discussed, we shall herewith attempt to evaluate the present state of the art in automatic indexing techniques, using such available criteria as seem most appropriate. First, we suggest that all of out initial questions except possibly the last, can today be answered affirmatively. "Is indexing by machine possible at all?" To this we can answer an unequivocal "yes" in view of the many examples of KWIC type indexes extant and in practical use. Secondly, "Is what can be done by machine properly termed 'abstracting', 'indexing', or 'classifying'?" If, by definition, word indexing of any kind is not "properly termed... indexing", then, as we have seen, automatic derivative indexing, such as KWIC, or the selection of words to serve as index tags based upon the frequencies of their occurrence in text, is not so either.

The fundamental Luhn concept for indexing based on word frequencies is, as we have seen, straightforward: namely that, after disregarding the most frequent "common words", especially those that are syntactic-function words -- articles, conjunctions, prepositions, and the like, together with those words that occur infrequently in a given text, the remaining high frequency words should give a reasonable indication of what the author was writing "about". Critiques of the Luhn position have been made on several-fold grounds:

(1) Information-theoretic - that, in fact, the most information is conveyed by the least frequent words.
(2) Absolute vs. relative frequencies of usage within specialized fields.
(3) Modifications of semantic purport by contextual and syntactic associations.
(4) Problems of synonymity and, conversely, of orthographically identical words. 1/
(5) Multi-aspect points of interest, and future need of access to material the author himself did not emphasize.

The last point raises again the criticisms that have been made against derivative, extractive or "word" indexing of all types. To repeat, although such procedures may index "as the author himself indexed best -- in his own language", the significant points are (1) there may be peripheral, minor, or unrecognized aspects of his topic and incidental information disclosed, of future interest to others, which the author himself is in no special position to recognize, and (2) notwithstanding the "author's own terminology" being current usage rather than the "fossilized" vocabulary of any previously established classification or indexing scheme, this very "currency" changes from field to field and, quite literally, from day to day. Nevertheless, it should be re-emphasized that the validity of these criticisms is not limited to automatic derivative indexing as such, but rather is applicable against any indexing system whatsoever, manual or machine, which is so strictly limited to author-terminology, author-emphases, and the consideration of the document at hand as a self-contained entity, without regard to other documents in a collection, in a particular field, and without respect to specific user needs. By contrast to this type of limitation, more promising approaches should stress both similarities and differences between a new document and previously received documents, between documents "belonging" to some definable category, or not, and even, as responsive to a particular user's profile-of-interest, or not.

1/ See Baxendale, 1962 [42], pp. 67-68: "... resolution of orthographic ambiguities is a non-trivial and over-riding prerequisite for the computer processing of text...", p. 67.

Derivative indexing, whether by man or machine, is thus subject to many disadvantages. First and foremost, it is constrained by a particular individual's personal manner of expression of concepts in language. This limitation is controlled only by his presumptive desire to communicate with some particular (more or less general, or more or less specialized) audience. His choices of natural language expressions, however, will be conditioned by at least some of the following factors:

(1) The range and precision of his personal mastery of both general and specialized vocabularies for a given time, place, and specialized field of discourse.

(2) His personal expectations as to the probable reactions (in the sense of effective communication) of his intended audience to the expressions that he does choose, involving all of the problems of different usages of technical terminology from field to field, from formal to informal presentations, from scholarly reviews to progress reports heavy in current "technese" and "fashionable words".

(3) His habits of thought and his training in his field.

(4) His awareness of more than one possible audience and of more than one point or topic of potential interest to his readers.

Secondly, indexing by the author's own words is remarkably sensitive to a particular period of time, so that the terminology becomes rapidly outdated and often seriously misleading in its connotations. Thirdly, the user has no advance knowledge of the terminology that has been used in all the varied texts of a collection and he must therefore be able to predict a wide variety of possible ways of expressing ideas in words, phrases, and even by implication. Fourthly, for collections indexed on a word-derivative basis, there is little or no possibility for generic searching. 1/ Finally, there is the more general question, applicable to both derivative and assignment indexing, of how well, ever, can a condensed representation serve the purposes of specific subject content recapture? In the strict sense, only by the elimination of truly redundant information. But even this is a relative matter. What is redundant for an author may not be so for several different potential users of the reports or papers that this author writes. What is redundant for one user is not necessarily so for others.

The further problem for machine techniques is therefore: how selection rules can be provided that will replicate a given human pattern of selectivity, or, alternatively, how selection rules can be established and defined that will produce an equivalent and comparable result - that is, one which typical users would agree is as pertinent to their query-answer relevance decisions as any available alternative.

Certainly the problem of appropriate selection is at the heart of the matter. This is a crucial question, even if we sort out and can specify the different uses, for a particular collection, a particular clientele, at a particular time, that automatically generated condensed document representations may have. Wyllys, in appraising automatic abstracting efforts, considers that the goal should be to provide extracts which will serve a search-tool function -- that is, they will furnish the searcher with enough information about the document content so that he may decide whether it is probably pertinent to his then interests or not and hence decide whether or not to read the document in full. By contrast, he says of the "content-revelatory function" that an abstract should: "furnish the reader with enough information about the related document so that in most cases he will not need to read it itself." 2/

1/ See for example, Doyle, 1963 [162], with respect to lack of capacity for generic searching as one of the major disadvantages of natural text search systems.

2/ Wyllys, 1963 [653], p. 6.

Let us recall the objections to the use of the terms "auto-encoding" (or "auto-indexing" or "auto-abstracting") because of the possible connotation of self-encoding, etc.. [1] This is an objection based upon avoiding ambiguous or misleading terminology, but it also points to an objection as to the principle involved--that is, of treating the document itself, in its own right, as a self-sufficient, self-contained, universe of discourse, and of assuming that some type of summation-condensation over a number of different and individually-derived representations of the separate documents in a collection can provide an effective selection-retrieval guidance system to the contents of various specific documents in that collection. Even when the actual operations are to be abetted by synonym reduction and normalization procedures (whether at the indexing or search negotiation stage, or both), there is a significant difference between this endogenous hypothesis and its exogenous alternative: that the basis for automatic indexing be the consensus of the collection, or of a sample of the collection, or of prior indexing.

Assignment indexing, especially in the sense that concept-indexing is the goal, may be subjectively preferable to derivative indexing not only because it involves exogenous emphases but because it tends to delimit, centralize, and standardize the access points available to the user in his search-retrieval operations. However, in terms of the human indexing situation, it involves all the traditional difficulties of indexing - which in turn invoke the problems of evaluating indexing systems:

> **"Justification for any indexing technique must ultimately be based on successful retrieval. Success can only be evaluated in terms of a closed system; that is, a system wherein sufficient knowledge is available of the entire contents of the materials, so that an evaluation can be made of various techniques as to their retrieval effectiveness. The various systems ... cannot really be weighed except on the basis of a test comparing one against the other. This has not been done in any place." [2]**

Nevertheless, there are a variety of reasons for accepting even the relatively crude derivative indexing products as practical tools today, for seeking machine-usable rules for the improvement of these products, and for continuing research efforts in automatic assignment indexing and automatic classification. There are, first and foremost, the cases where conventional indexes are inadequate or non-existent. Thus Wyllys claims:

> **"It is well-known that the current methods of producing, through human efforts, condensed representations of documents are already hopelessly inadequate to cope with the present volume of scientific and technical literature. Many papers are never indexed or abstracted at all, and even in the cases of those that are indexed or abstracted, the indexes and abstracts do not become available until six months to two years after the publication of the paper." [3]**

Again, with respect to automatic derivative indexing, especially KWIC indexes based on titles alone, there can be no question as to the evaluation criterion of timeliness. The success of this aspect is widely acknowledged by users, systems planners, and interested observers. On the other hand, there is very little reported evidence available on which

---

[1] See p. 3 of this report.

[2] Black, 1963 [64], p. 16.

[3] Wyllys, 1961 [650], p. 6.

any objective measure of comparative cost-benefit ratios may be obtained. Black reports, but without supporting data, that:

> "It has been estimated that the efficiency of KWIC indexing is about 76 per cent compared with about 82 per cent for conventional indexing or classification." 1/

White and Walsh report that:

> "From the limited experiment on methods of indexing the 1962 issues of the Abstracts of Computer Literature, the permuted title indexing retrieved only 52 percent of the information. This low percentage may be attributed to the changing and not yet uniformly standardized terminology existing in computer technology." 2/

KWIC indexes, because of their very currency, are fulfilling significant maintaining-awareness needs today. Improved titling practice, enforced by editorial rigor or contractual requirements or both, can improve their usefulness. They fill gaps in the bench scientist's or engineer's ability to know about what might be of interest to him, either because the material is not otherwise covered in normal secondary publication (e.g., conferences and proceedings of symposia, internal technical reports not produced on Government contracts and therefore not announced and indexed by the cognizant agencies, and the like) or because the sheer bulk of the product of indexing-abstracting services in his field prevents his effective use of these services unless more specific access points are provided. The claim that "something is better than nothing" is not without merit, 3/ even with all the problems of non-resolution of synonymity, homography, topical scatter, long blocks of entries under the sorting term, the even more significant disadvantages of author-bias towards his principle topic, the author's choice both of emphasis and terminology, and the like. Williams, considering word-with-context indexes, whether limited to title only or to titles with readily available augmentation, makes the following comments:

> "Limitations and other troublesome features of the method have been obvious, but perhaps over obvious, in the light of its growing acceptance and of the basic validity of permitting a document to speak for itself, even in a much abstracted recapitulation. Wherever there are large and growing problems in maintaining publication schedules for established subject indexes, or wherever pressing needs develop for more frequent indexes, for rapid, low-cost cumulation, or for indexes in areas where suitable indexing services are wanting, there no apology is needed for proposing that this method be considered and tried, as a precursor to 'better' indexing, if not as a substitute. Its use may be of interest also in less troubled circumstances, in its own right, and because of common elements involved in its production and the provision of other wanted products and functions (catalog records, current-awareness, lists, etc)." 4/

Returning to the question of whether automatic indexing is possible, it can be seen that, at least in the derivative indexing sense, it is not only possible but can be practically useful. To dismiss the evidence of automatic derivative indexing operations that are in production today by rigorous definition of what indexing is in effect anticipates both our

---

1/ Black, 1962 [65], p. 318.

2/ White and Walsh, 1963 [639], p. 346.

3/ See Veilleux, 1962 [624], p. 81: "Accepting the premise that partial control of information satisfies more consumers than absence of control, perfection was traded for currency."

4/ T.M. Williams, private communication, dated January 4, 1962.

third and fourth questions: whether machine-generated indexes are as good or better than the products of human operations and of how we can measure and appraise the adequacy of any indexing system whatever. Here are encountered the "core" problems of meaning in communication, of information loss in any reductive transformation of actual messages or documents, of relevance of particular messages to particular queries and to particular human needs, of judgments of relevance.

Because of these underlying yet overriding questions, the state-of-the-art in the evaluation of indexing systems is in fact far more primitive than that of automatic indexing itself. An easy, and an early, solution is not likely. Therefore, today, in appraising machine potentials for assignment indexing we are faced with what is in effect a single criterion: namely, will a given group of human evaluators, whatever their standards and requirements, agree as much with the products of an automatic indexing procedure, otherwise competitive on a cost-benefit ratio with human indexing of the same material, as they do amongst themselves?

Within the limits of small, specially selected samples of document or message collections, it is possible to demonstrate that:

(1) Replication of the products of at least some existing systems, within the consistency levels observed for these systems, can be achieved.

(2) Retrieval effectiveness with respect to relevant items indexed by automatic assignment procedures can be at least as good as, and may be superior to, that obtained from run-of-the-mill manual indexing of the same items.

(3) Costs of indexing can be held at or below the costs of equivalent manual indexing, provided both that the input material required is already in machine-usable form, or can be held to an average of, say, 100 words or less, and that the clue-word lists, association factors, or probabilistic calculations can be accommodated within internal memory.

(4) Significant gains in time required to generate an index or to index or re-index a collection can be achieved.

Some degree of theoretical success in assignment indexing by machine can thus certainly be claimed. Moreover, many of the test results reported do clearly indicate a quality of indexing, for a given collection at a given level of specificity of indexing, at least comparable to that which is typically and routinely achieved by people in a practical indexing situation. No more should be asked of the automatic techniques unless better human indexing can be specified as being equally feasible, timely, and practical. Further, no more should be asked of automatic techniques in terms of the evaluation of their potentialities, than is now asked of the manually-prepared alternatives. 1/

Data with respect to comparison of the results of automatic assignment indexing techniques to either a priori or a posteriori human judgment have been mentioned previously in this report in terms of actual test results reported, and the most significant of these reported data are summarized in Table 2. 2/ Typically, however, these data reflect, in varying degrees, so small a sample of test cases, of user preferences, and/or of special purpose and interest, that no general extropolation is reasonable. Moreover, the general questions of the "core" problems of evaluation in general again rear their own ugly heads.

1/ Compare, for example, Kennedy, 1962 [311] and Needham, 1963 [433].

2/ See pp. 101-103 of this report.

Thus, Borko and Bernick point out:

> "Up to this point we have used human classification as our criterion for the accuracy of automatic document classification. Against this criterion we have been able to predict with approximately 55% accuracy, and no more. Is this because out techniques of automatic classification are not very good, or is it because our criterion of human classification is not very reliable? There is some evidence to indicate that the reliability of human indexers is not very high. The reliability of classifying technical reports needs investigating and, perhaps even more basically, the reasons for using human classification as a criterion at all." 1/

In general, the results of automatic index-term assignment procedures appear to run in the area of 45-75 percent agreement with prior human indexing, 2/ and this in turn is well within range of, and often superior to, estimates of human inter-indexer consistency based on actual observations and tests. There can be little or no doubt that the results of automatic assignment indexing experiments to date, (if extrapolation from the small and often highly specialized samples so far used in actual tests is in fact warranted 3/) do suggest that an indexing quality generally comparable to that achievable by run-of-the-mill manual operations, at comparable costs and with increased timeliness, can be achieved by machine.

The question which remains is simply that of practicality, today. Extrapolation from small samples is highly dangerous, as is well noted even by enthusiastis for machine techniques. The fact that for at least some systems, the limitations on number of clue words that can be handled (due in part to computational requirements, matrix manipulations, and the like) are such that, even in an experimental situation, certain "tests" are excluded from the result statistics, because the items contained an insufficient number of clues, is a serious indictment of reasonable extrapolations for these techniques today. Most tests so far reported have involved not only a highly specialized "sample" library or collection, but a severe limitation on the total number of "descriptors", subject headings, or classification categories to be assigned. Maron used 32, Borko 21, Williams 20, SADSACT 70, Swanson 24. How would any of these approaches fare, given several hundred, much less

---

1/ Borko and Bernick, 1963 [78], pp. 31-32.

2/ See Table 2.

3/ This is an important, perhaps crucial, caveat. See, for example, Goldwyn, 1963 [233], p. 321: "In the micro-experiments of many of those who would apply statistical techniques ... The document collection consists of 0-100 units. Results based on the manipulation, real or imagined, of such a collection can be valid for it, yet become shaky or even nonapplicable to larger collections"; Perry 1958 [471], p. 415: "A degree of selectivity quite acceptable for files of moderate size may prove quite inadequate in dealing with large files. This fact often makes it necessary to exert unusual care and considerable reserve in evaluating the results of small-scale tests and demonstrations which may tend to cause the mass effects of large files to be underestimated or overlooked completely"; Swanson, 1962 [586], p. 288: "The extent to which semantic characteristics of natural language are susceptible to being generalized from small sample data is deceptive."

several thousand, possible indexing or classificatory labels? 1/

The use of very brief short articles, or of abstracts, as the members of experimental corpora for investigations of automatic assignment indexing techniques presuming the processing of full text, either for indexing purposes or for subsequent "indexing-at-time-of search", is seriously misleading. First, it is not truly representative of discursive text, either in vocabulary-syntax, or stylistic variations involving synonymity, tropes, elisions, dangling referents, and inumerable other meaning-implications, not explicitly stated.

Secondly, as any author of a technical paper, for which he must provide an abstract, knows all too well, he must concentrate in the abstract on a telegraphic emphasis toward his principal topic and the points he wishes to make. He must omit most qualifying, specifying, and suggestive-of-other-leads-or-applications words and phrases, which he will in fact develop in the text itself. For this reason, even supposing that the author himself is unusually well-aware of the multiple points of access that many different potential users might desire, the required brevity of the abstract form almost necessarily demands terse, shorthand-type statements that can only increase the problems of "technese", of homography, and of single-subject representation.

Granted, in either manual or machine-serviceable systems today, the current-awareness scanning need is largely met by indexing based solely or primarily on title only, or title-plus-abstract. But is this good enough for search and retrieval? If and only if it is, then automatic indexing potentialities available today should be considered for both purposes.

Our final question as to whether automatic indexing can be accomplished by statistical means alone or must involve syntactic, semantic and pragmatic considerations is not entirely answerable. In terms of achieving comparable quality with many manually prepared indexes available today, statistical means alone do appear promising. But is the achievement of just this level (even if accompanied by significant gains in timeliness, coverage, and economy) really good enough? There are a number of serious investigators

---

1/ For example, Black predicts (1963) [64], p. 19) that for most systems an adequate vocabulary or thesaurus will comprise some twenty thousand terms. See also Arthur D. Little, Inc., 1963 [23], p. 65: "The enormous number of computations required increases very rapidly with the number of indexing terms. Existing computers, operating serially, do not appear to be capable of handling the problem economically for collections with 9000 or more terms even if the simplest associative techniques are employed"; Williams, 1963 [642], p. 162: "One of the practical problems... is in the inversion of large matrices. In certain methods the order of the matrix will equal the number of different word types in the population, which is usually in the thousands."

convinced that it is not, 1/ and for this reason, research efforts are being directed toward these other considerations.

On-going research and development work - whether in modified derivative indexing approaching a "concept-indexing" level; in automatic assignment indexing techniques as such; in automatic classification or categorization procedures, or in potentially related efforts directed toward automatic abstracting, automatic content analysis, and other aspects of linguistic data processing - is both reasonably extensive and quite promising. Most of the investigators who are seriously active in the field report their current objectives and recent accomplishments regularly to the National Science Foundation for publication in the series "Current Research and Development Efforts in Scientific Documentation." In the most recent issue, unfortunately current only as of November, 1962, there are not less than 25 reports of KWIC and similar title-permuted derivative indexing methods generated or proposed-to-be-generated by machine, there are several instances of investigations into various possibilities of modified derivative indexing to be accomplished by machine, and there are five to ten reports of active experimentation with various automatic assignment indexing schemes. These efforts and even more recently organized projects point in the hopeful direction that "KWIC indexes should be merely a sample of things to come". 2/

Assignment indexing techniques so far investigated can be, as we have seen, of two types which are quite distinct in terms of the principles involved. The first, which can be the more readily mechanized, involves the use of thesaurus-type lookup procedures covering the definable rules of "scope notes", "authority lists", or "see also" reference practice. The second type of assignment indexing, however, depends upon decision-making as to the propriety of assigning a particular indexing term to a particular document with reference to assignments to the collection as a whole (or a sample thereof). This latter type of assignment may be in terms of *a priori* categorizations of separable subsets of the collection.

Alternatively, the bases for the latter type assignment-indexing procedures may be derived from *a posteriori* determinations of the suitable subsets as in the factor analysis experiments of Borko, the latent class analysis approach of Baker, and the clustering-clumping approaches to automatic classification of Needham and others. It is to be noted in particular that Needham thinks an automatically generated categorization is preferable precisely because of lack of knowledge as to the exact attributes defining a class in

---

1/ See, for example, Climenson et al, 1962 [133], p. 178: "The statistical approach attempts to use no more than the occurrences of word spellings and their relative distances in the document environment ... [and] cannot provide the discrimination necessary for most indexing and abstracting applications"; Doyle, 1963 [162], p. 3: "Automatic indexing and abstracting, as currently conceived, do not require any sort of dictionary or other semantic reference, but only counting, comparing, and sorting-operations well known in numerical data processing. But success in applying such rules on a purely automatic basis can't help but be limited"; Borko, 1962 [75], p. 5: "Although difficult, identification [of different meanings carried by the same word, of the same meaning carried by different words] must be accomplished before the automatic categorization of document content can be truly effective. For the most part statistical methods, and even syntactic analysis, are inadequate for the job. A technique of textual analysis based upon the semantic properties of language is needed"; Grosch, 1959 [244], p. 20: "We need semantic methods ... that will look for the intersection of redundant descriptors, each of which is at least slightly erroneous."

2/ Doyle, 1962 [163], p. 381.

existing classification schemes. However, in the related field of pattern recognition Uhr and Vossler have shown promising results both for criterial feature analysis (a priori assumption as to attributes or properties governing membership in specified classes) and for randomly generated discrimination operators which, applied in a recursive manner, are increasingly adaptive to the detection of class-mambership (Uhr and Vossler, 1961 [615]).

One particular way of looking at the problems of automatic indexing results, in effect, in placing these problems within the broader field of pattern perception and pattern recognition. We suggest that this is in fact a particularly fruitful approach. Certainly there is a wide area of potential commonality, and many promising leads for further research in automatic categorization can be found in the general pattern recognition literature, especially in work on randomly generated operators and on the problems of determination of membership in classes. 1/ Conversely, automatic classification techniques originally conceived as applicable to the handling of documentary information have in fact been applied quite successfully to at least one case of groupings of physical objects on the bases of machine-detectable common properties.

The question of determination of membership-in-classes is basic to the problems of automatic classification and categorization. Thus the techniques for discriminating the statistically significant associations between "properties" of objects or items that are to be grouped into classes or categories, even when such "properties" are not known in advance and have no a priori identification, point to an increasing and promising convergence of research in pattern recognition, propaganda analysis and psycholinguistics, mathematics and statistics, studies of linear threshold devices, and the like, as well as in the linguistic data processing field as such.

It is true that such synthesized "classes" may have no convenient "names" or linguistic interpretations which make much sense to the individual human searcher or user. Nevertheless, what is suggested is that a radical departure from conventional habits of literature search and retrieval may be desirable from the standpoint of effective use of machine potentialities. This might mean that, ab initio, the customer would pose to the system a search query request not couched in his notion of words or terms actually used in the system, but either (a) an outline or statement of his own research proposal and plan of attack or (b) an indication of one or several items that he has already decided are pertinent to his interests, with a request for "more like these".

An equally radical departure from conventional present habits and thinking is already implicit in Needham's suggestion of an automatically derived classification system and manual assignments thereto. 2/ It would attack present-day machine capacity and processing time limitations such that property and class or category associations must be held to something less than 1,000 x 1,000, unless prohibitive processing costs are to be incurred. This approach would assume a one-time large-scale building of vocabulary and term or category associations and derivation of assignment algorithms, and the printing out of the results in multiple copies for use by low-level clerical personnel carrying out, indeed, "machine-like" indexing.

A final promising approach to the future prospects for fully automatic indexing and categorization is the perseverance in research and development efforts in advance of the

---

1/ See, for example, Sebesyten, 1961 [539], 1962 [538].

2/ Needham, 1963 [432], p. 1.

advent of versatile character readers and inexpensive, very large capacity, rapid direct access memories. These efforts will include not only further systematic exploration of syntactic, semantic and pragmatic considerations in linguistic data processing, but also further attacks on the problems of language and meaning themselves. Thus, we may conclude with Maron that: "automatic indexing represents the opening wedge in a general attack at not only the problems of identification search and retrieval, but also the problem of automatically transforming information on the basis of its content." 1/

If we are to attempt to solve this problem, as indeed we should, must we not look forward to the possibilities of rapid up-dating, thesaurus growth and revision, and quick and economical re-indexings of entire collections that only machine-processing capabilities can promise today?

ACKNOWLEDGEMENTS

The contributions of Miss Josephine L. Walkowicz and her staff in the preparation and checking of items for the bibliography, and of Mrs. Betty J. Anderson, Mrs. Helen B. Grantham, and Mrs. Anna K. Smilow in the typing and editing of the manuscript are gratefully acknowledged. The courtesy of Miss Thyllis Williams, Mr. Joseph Becker, Mr. Herbert Ohlman, and the late Hans Peter Luhn in making available unpublished materials is also gratefully acknowledged.

1/ **Maron, 1961, [395], p. 240. See also Salton, 1962 [518], p. 234 and Borko and Bernick, 1962 [77], p. 3**

APPENDIX A: LIST OF REFERENCES CITED AND SELECTED BIBLIOGRAPHY

1. "Actes du Colloque sur le Mechanisation de Recherches Lexicologiques", (Besançon, June 6-10, 1961), Les Cahiers de Lexicologie 3, 1-220 (1961).

2. Adair, W.C. "Citation Indexes for Scientific Literature?", Amer. Documentation 6, 31-32 (1955).

3. Allen, G., L. Cavalli-Sforza, J. Lederberg, G. LeFevre, J. Melnick and S. Spiegelman, "Research and Evaluation Program on Citation Indexing", Institute for Scientific Information, Philadelphia, Pa. 19 Oct 1962.

4. Alvord, D. "King County Public Library Does it with IBM", Pacific Northwest Library Assoc. Q. 16, 123-132 (1952).

5. American Diabetes Association, "Diabetes-Related Literature Index by Authors and Key Words in the Title for the Year 1960", Vol 12, Suppl. 1 of Diabetes, The Journal of the American Diabetes Association, (1963).

6. (American Federation of Information Processing Societies)*, "Proceedings of the Western Joint Computer Conference, 1959", Vol 15, Institute of Radio Engineers, New York, 1959, 360 p.

7. (American Federation of Information Processing Societies)*, "Proceedings of the Western Joint Computer Conference 1961, Extending Man's Intellect", Vol 19, Western Joint Computer Conference, Glendale, Cal. 1961, 661 p.

8. American Federation of Information Processing Societies, "Proceedings of the Spring Joint Computer Conference, 1962", Vol 21, National Press, Palo Alto, Cal. 1962, 314 p.

9. American Federation of Information Processing Societies, "Fall Joint Computer Conference, 1962", AFIPS Conference Proceedings, Vol 22, Spartan Books, Washington, D.C. 1962, 314 p.

10. American Federation of Information Processing Societies, "Fall Joint Computer Conference, 1963", AFIPS Conference Proceedings, Vol 24, Spartan Books, Baltimore, Md. 1963, 647 p.

11. American Meteorological Society, "Examples of Keyword - U.D.C. Indexes Compiled on Electronic Computer (IBM 704) and Tabulator (IBM 407) From Contents of Periodicals and Serials Listed in MGA", Meteorological and Geoastrophysical Abstracts, XII (Mar 1961, Nov 1961).

12. American Meteorological Society, "Meteorological and Astrophysical Titles", Vol 1, no. 1, Washington, D.C. Apr 1961. Vol 1, no. 2, Oct 1961.

13. American Meteorological Society, "Meteorological and Geoastrophysical Titles", 2:1 (1962). (Second experimental issue) Washington, D.C., 55 p.

14. The American University, "Machine Indexing: Progress and Problems". (Papers presented at the Third Institute on Information Storage and Retrieval, Feb 13-17, 1961). Washington, D.C. 1962, 354 p.

* Note that although proceedings of the Joint Computer Conferences were not published by the American Federation of Information Processing Societies prior to Volume 20, they are here grouped in accordance with the volume series numbers.

15. Anger, A. "A Class of Reference-Providing Information Retrieval Systems", in G. Salton [ed]. "Information Storage and Retrieval, No. 1", 30 Nov 1961, p. III-1 to III-30.

16. Anzlowar, B.R. "Abstract Automation in Drug Documentation", in H.P. Luhn [ed]. "Automation and Scientific Communication, Short Papers, Pt. 1", 1963, p. 103-104.

17. Armed Services Technical Information Agency, "Controlling Literature by Automation". (Presented at the IV Annual Military Librarians Workshop Sponsored by Armed Forces Technical Information Agency, 5-7 Oct 1960.) Washington, D.C., 1960, 130 p.

18. Armed Services Technical Information Agency, "Key-Words-In-Context Title Index. A List of Titles for ASTIA Documents Not Previously Announced", No. 1, Arlington, Va. Oct 1962, 156 p.

19. Armed Services Technical Information Agency, "Key-Words-In-Context Title Index", No. 2., Arlington, Va. Feb 1963, 117 p.

20. Artandi, S. "Book Indexing by Computer", Doctoral Dissertation, Rutgers University Graduate School of Library Science, Mar 1963, 207 p., available through University Microfilms, Inc., Ann Arbor, Mich. 1963.

21. Artandi, S. "A Selected Bibliographic Survey of Automatic Indexing Methods", Spec. Libraries 54, 630-634 (1963).

22. Artandi, S. "Thesaurus Controls Automatic Book Indexing by Computer", in H.P. Luhn [ed]. "Automation and Scientific Communication, Short Papers, Pt. 1", 1963, p. 1-2.

23. Arthur D. Little, Inc. "Centralization and Documentation", Final report to the National Science Foundation, C-64469, Cambridge, Mass. July 1963, 70 p.

24. Asher, J.W. and M. Kurfeerst, "The High Speed Computer as a Research and Operations Device in School Law", Cooperative Research Project No. 1275, School of Education, University of Pittsburgh, Pittsburgh, Pa. Feb 1963, 66 p.

25. Atherton, P. "A Collection of Remarks about Citation Indexes", American Institute of Physics, New York, Apr 1962, 6 p.

26. Atherton, P. and J.C. Yovich, "Three Experiments with Citation Indexing and Bibliographic Coupling of Physics Literature", American Institute of Physics, New York, Apr 1962, 39 p.

27. Baker, F.B. "Information Retrieval Based on Latent Class Analysis", J. Assoc. Computing Machinery 9, 512-521 (1962).

28. Balz, C.F. and R.H. Stanwood, "Literature Dissemination and Retrieval Using the Merge System", in H.P. Luhn [ed]. "Automation and Scientific Communication, Short Papers, Pt. 1", 1963, p. 61-62.

29. Balz, C.F. and R.H. Stanwood, "Literature on Information Retrieval and Machine Translation", International Business Machines Corp. Owego, N.Y. Nov 1962, 117 p.

30. Balz, C.F. and R.H. Stanwood, "On Preparing Information for KWIC Indexing (IBM 7090)", Rept. No. 62-816-729, International Business Machines Corp. Owego, N.Y. 15 Jan 1962, 36 p.

31. Balz, C.F. and R.H. Stanwood, "Some Applications of the KWIC Indexing System", Rept. No. 62-825-475, International Business Machines Corp. Owego, N.Y. 15 June 1962, 12 p.

32. Bar-Hillel, Y. "A Logician's Reaction to Recent Theorizing on Information Search Systems", Amer. Documentation 8, 103-113 (1957).

33. Bar Hillel, Y. "The Mechanization of Literature Searching", in National Physical Laboratory, "Mechanization of Thought Processes", Symposium No. 10, Vol II, 1959, p. 791-807.

34. Bar-Hillel, Y. "Some Theoretical Aspects of the Mechanization of Literature Searching", Tech. Rept. no. 3, Hebrew University, Jerusalem, Apr 1960, 74 p.

35. Bar-Hillel, Y. "Theoretical Aspects of the Mechanization of Literature Searching", in W. Hoffman [ed]. "Digital Information Processors", 1962, p. 406-443.

36. Barnes, A.B. and A. Resnick, "The Effect of Varying Word Lengths on the Accuracy of Matching Documents with Reader's Interest", preprint of paper presented at the ACM 1963 National Conference, Denver, Colo. Aug 1963.

37. Barnes, R.F. "Language Problems Posed by Heavily Structured Data", Comm. Assoc. Computing Machinery 5, 28-34 (1962).

38. Barnes, R.F. "Lectures on Modern Logic and Automatic Document Analysis", presented at the NATO Advanced Study Institute on Automatic Document Analysis, Venice, 7-20 July 1963.

39. Baxendale, P.B. "Automatic Processing for a Limited Type of Document Retrieval System", in H.P. Luhn [ed]. "Automation and Scientific Communication, Short Papers, Pt. 1", 1963, p. 67-68.

40. Baxendale, P.B. "An Empirical Model for Computer Indexing", in "Machine Indexing", American U., 1962, p. 207-218.

41. Baxendale, P.B. "Machine-Made Index for Technical Literature--an Experiment", IBM J. Research and Development 2, 354-361 (1958).

42. Baxendale, P.B. "Man-Computer Indexing: Functions, Goals, and Realizations", in "Joint Man-Computer Indexing and Abstracting", Mitre SS-13, 1962, p. 61-73.

43. Becker, J. "Present and Future Applications of International Business Machines to Libraries", unpublished paper, presented at Catholic University, Washington, D.C. 1947, 15 p.

44. Becker, J. "Some Approaches to Mechanization of Technical Information Processing Systems", in "Proceedings of the March AFBMD Conference", 1960, p. 9-20.

45. Becker, J. and R.M. Hayes, "Information Storage and Retrieval: Tools, Elements, Theories", Wiley, New York, 1963, 448 p.

46. Bell Telephone Laboratories, Inc. "BTL Talks and Papers 1962". (First of an annual series) Murray Hill, N.J. 1963, lv.

47. Bell Telephone Laboratories, Inc. "Index to the Literature of Magnetism", Vol 2, 1961-1962, Murray Hill, N.J. 193 p.

48. Bell Telephone Laboratories, Inc. "Mechanized Indexing of Internal Reports", Murray Hill, N.J. Jan 1961, 18 p.

49. Bennett, E. and J. Spiegel, "Document and Message Routing Through Communication Content Analysis" in M.L. Juncosa [ed]. "Symposium on Optimum Routing in Large Networks", 1962, p. 718-719.

50. Bennett, J.L. "A System for Transcribing Printed Text into a Machine Readable Format", in H.P. Luhn [ed]. "Automation and Scientific Communication, Short Papers, Pt. 1", 1963, p. 141-142.

51. Berg, R.M. "Future Plans for Mechanization", in "The Literature of Nuclear Science: Its Management and Use", U.S. Atomic Energy Commission, Dec 1962, p. 201-204.

52. Bernard, J. and C.W. Shilling, "Accuracy of Titles in Describing Content of Biological Sciences Articles", BSCP Communique 10-63, Biological Abstracts, Philadelphia, Pa. May 1963.

53. Bernier, C.L. "Correlative Indexes I. Alphabetical Correlative Indexes", Amer. Documentation 7, 283-288 (1956).

54. Bernier, C.L. "Language and Indexes", Amer. Documentation 7, 222-224 (1956). Also in J.H. Shera et al [eds]. "Documentation in Action", 1956, p. 325-329.

55. Bernier, C.L. "Organizing Abstract Information", (unpublished paper presented at the American Documentation Institute, 6 Nov 1953, cited by C.L. Bernier, "Correlative Indexes I", p. 284 and by M. Taube et al, "Studies in Coordinate Indexing, II", p. 73.)

56. Bernier, C.L. and E.J. Crane, "Correlative Indexes VIII: Subject-Indexing vs. Word-Indexing", J. Chem. Documentation 2, 117-122 (1962).

57. Bernier, C.L. and K.F. Heumann, "Correlative Indexes III. Semantic Relations Among Semantemes--The Technical Thesaurus", Amer. Documentation 8, 211-220 (1957).

58. Berry, M.M. "Application of Punched Cards to Library Routines", in R.F. Casey, et al, "Punched Cards: Their Applications to Science and Industry", 1958, p. 279-302.

59. Bessinger, J.B. "Computer Techniques for an Old English Concordance", Amer. Documentation 12, 227-229 (1961).

60. Biological Abstracts, Inc. "Accuracy of Titles in Describing Content of Biological Sciences Abstracts", BSCP Communique, 15-63, Philadelphia, Pa. Sept 1963.

61. Biological Abstracts, Inc. "B.A.S.I.C. (Biological Abstracts' Subjects in Context)" 39:2 (15 July 1962) Philadelphia, Pa. 109 p. Issued semi-monthly.

62. Biological Abstracts, Inc. "Biochemical Title Index", Vol 1, no. 1, Philadelphia, Pa. Jan 1962. (This first issue contains a B.A.S.I.C. index).

63. Biological Abstracts, Inc. "Biological Abstracts", Vol 36, no. 20, Philadelphia, Pa. Oct 1961. (First appearance of the B.A.S.I.C. index).

64. Black, D.V. "Indexing Techniques Description and Background", Appendix to "Document Storage and Retrieval Techniques", Planning Research Corp., Los Angeles, Cal. 13 June 1963, 29 p.

65. Black, J.D. "The Keyword: Its Use in Abstracting, Indexing and Retrieving Information", ASLIB Proc. 14, 313-321 (1962).

66. Blackwell, F.W. "ALMS Analytic Language Manipulation System". Presented at the ACM 1963 National Conference, Denver, Colo. Aug 1963.

67. Boaz, M. [ed]. "Modern Trends in Documentation", proceedings of a Symposium held at the University of Southern California Apr 1958, Pergamon Press, New York, 1959, 103 p.

68. Bobrow, D.G. "Syntactic Analysis of English by Computer - A Survey", in "Proceedings of the Fall Joint Computer Conference, 1963", p. 365-387.

69. Bohnert, L.M. "New Role of Machines in Document Retrieval: Definitions and Scope", in "Machine Indexing", American U. 1962, p. 8-21.

70. Booth, A., L. Brandwood, and J. P. Cleave, "Mechanical Resolution of Linguistic Problems", Academic Press, New York, 1963, 306 p.

71. Borko, H. "Automatic Document Classification Using a Mathematically Derived Classification System", FN-6164, System Development Corp., Santa Monica, Cal. 28 Dec 1961.

72. Borko, H. [ed]. "Computer Applications in the Behavioral Sciences", Prentice-Hall, Inc., Englewood Cliffs, N. J. 1962, 633 p.

73. Borko, H. "The Construction of an Empirically Based Mathematically Derived Classification System", Rept. No. SP-585, System Development Corp., Santa Monica, Cal. 26 Oct 1961, 23 p. Also in American Federation of Information Processing Societies, "Proceedings of the Spring Joint Computer Conference, 1962", p. 279-289.

74. Borko, H. "Evaluating the Effectiveness of Information Retrieval Systems", Rept. SP-909/000/00, System Development Corp., Santa Monica, Cal. 2 Aug 1962, 8 p.

75. Borko, H. "Information Retrieval and Linguistics Project", Prog. rept. Tech memo. No. TM-676, System Development Corp., Santa Monica, Cal. 29 Jan 1962, 10 p.

76. Borko, H. "Research in Document Classification and File Organization", Rept. no. SP-1423, System Development Corp., Santa Monica, Cal. 13 Nov 1963, 12 p.

77. Borko, H. and M. D. Bernick, "Automatic Document Classification", Tech. memo. TM-771, System Development Corp., Santa Monica, Cal. 15 Nov 1962, 19 p. Also in J. Assoc. Computing Machinery 10, 151-162 (1963).

78. Borko, H. and M. D. Bernick, "Automatic Document Classification, Part II- Additional Experiments", Tech. memo. TM-771/001/00, System Development Corp., Santa Monica, Cal. 18 Oct 1963, 33 p.

79. Borko, H. and M. D. Bernick, "Toward the Establishment of a Computer Based Classification System for Scientific Documentation", Rept. no. TM-1763, System Development Corp., Santa Monica, Cal. 19 Feb 1964, 47 p.

80. Brandenberg, W. "Write Titles for Machine Index Information Retreival Systems", in H. P. Luhn [ed]. "Automation and Scientific Communication, Short Papers, Pt. 1", 1963, p. 57-58.

81. Bristol, R. P. "Can Analysis of Information be Mechanized?", College and Research Libraries 13, 131-135 (1952).

82. Brownson, H. L. "New Developments in Information Storage and Retrieval", in C. Popplewell [ed]. "Information Processing 1962", 1963, p. 294-295.

83. Buckland, L. F. "Machine Recording of Textual Information During the Publication of Scientific Journals", report on work done on National Science Foundation Contract 305, Inforonics, Inc., Maynard, Mass. 16 Dec 1963.

84. Buckland, L. F. "Recording Text Information in Machine Form at the Time of Primary Publication", in H. P. Luhn [ed]. "Automation and Scientific Communication, Short Papers, Pt. 2", 1963, p. 309-310.

85. Busa, R. "Complete Index Verborum of St. Thomas Aq.", Speculum, A Journal of Mediaeval Studies, 424-425 (1950).

86. Busa, R. "Die Elektronentechnik in der Mechanisierung der Sprachwissenschaftlichen Analyse", Nach. für Dok. 7, 7 (1957).

87. Busa, R. "Entwicklungen der Mechanisierung der Sprachlichen Analyse", Nach. für Dok. 4, 202-204 (1953).

88. Busa, R. "The Index of All Non-Biblical Dead Sea Scrolls Published up to December 1957", Revue de Qumran 1, 187-197 (1958).

89. Busa, R. "Mechanisierung der Philologischen Analyse", Nach. für Dok. 3, 14-19 (1952).

90. Busa, R. "Sancti Thomae Aquinatis Hymnorum Ritualium. Varia Specimina Concordantiarum. Primo Saggio Di Parole Automaticamente Composti E Stampati Da Macchine IBM A Schede Perforate. (A First Example of a Word Index Automatically Compiled and Printed by IBM Punch Card Machines)". Fratelli Bocca, Milan, 1951, 180 p.

91. Busa, R. "Summary of the Experience of the Centro Per L'Automazione Dell Analisi Letteraria of the Aliosianum", paper presented at the Symposium on Machine Methods for Literary Analysis and Lexicography, 24-26 Nov 1960, Tübingen, Germany, Sep 1960, lv.

92. Busa, R. "The Use of Punched Cards in Linguistic Analysis", in R. W. Casey et al, [eds]. "Punched Cards: Their Applications to Science and Industry", 1958, p. 357-373.

93. Bush, V. "As we may think", The Atlantic Monthly 176, 101-108 (1945).

-- Bush Committee report. See U.S. Department of Commerce, "Report to the Secretary of Commerce by the Advisory Committee on the Application of Machines to Patent Office Problems", 1954.

94. Bushnell, D. and H. Borko, "Information Retrieval Systems and Education", Rept. No. SP-947/000/01, (presented at the American Psychological Association Convention, St. Louis, Mo.), 18 Sep 1962, System Development Corp., Santa Monica, Cal. 1 Sep 1962.

95. "California Concordance Program Available", The Finite String 1, 1-4 (1964).

96. Callander, T.E. "Machine Reproduction of Catalogue Entries", Library Assoc. Record 52, 115-118 (1950).

97. Callander, T.E. "Punched Card Systems: Their Application to Library Technique", Library Assoc. Record 48, 27-31 (1946).

98. Callander, T.E. "Punched Card Systems in the Public Library", Library Assoc. Conf. Papers, Brighton, 1947, p. 23-28.

99. Carlsen, R.D., W.H. Gerner and H.S. Marshall, "Information Control", Industrial Engineering Dept. 564, Ref. 11, Rocketdyne Div., North American Aviation, Canoga Park, Cal. 1 Aug. 1958.

100. Carlson, G. "Letter to the Editor", Amer. Documentation 14, 328-329 (1963).

101. Carlson, W.H. "The Holy Grail Evades the Search", Amer. Documentation 14, 207-212 (1963).

102. Carroll, K.D. and R.K. Summit, "MATICO: Machine Applications to Technical Information Center Operations", Rept. no. 5-13-62-1, Lockheed Missiles and Space Co., Sunnyvale,Cal. Sep 1962, 24 p.

103. Casey, R.S., J.W. Perry, M.M. Berry and A. Kent, "Punched Cards: Their Applications to Science and Industry", Reinhold Publishing Corp., New York, second edition, 1958, 697 p.

104. Cassotta, L., S. Feldstein and J. Jaffe, "AVTA: A Device for Automatic Vocal Transaction Analysis", J. Experimental Analysis Behavior 4, 99-104 (1964).

105. C.E.I.R. Inc. "Design and Implementation of a Processing System to Create a 'Catalog of Research Report Titles Indexed By Keywords and Corporate Author'", Final Report, Arlington, Va. 28 Sep 1962, 29 p.

106. Centre d'Etude du Vocabulaire Francais, "Specimens De Travaux Lexicographiques Et Lexicologiques Realises Par le Laboratorie D'Analyse Lexicologique (Examples of Lexicographical and Lexicological Work at the Laboratory of Lexicological Analysis)", Besançon, France, 1960, 52 p.

107. Cezairliyan, A.O., P.S. Lykoudis and Y.S. Touloukian, "A New Method for the Search of Scientific Literature Through Abstracting Journals", J. Chem Documentation 2, 86-92 (1962).

108. Chasen, L.I. "Planning, Organizing and Implementing Mechanized Systems in a Space Technology Library", in H.P. Luhn [ed]. "Automation and Scientific Communication, Short Papers, Pt. 2", 1963, p. 303-305.

109. The Chemical Abstracts Service, "Chemical Biological Activities", sample issue, Columbus, Ohio, Sep 1962, 82 p.

110. The Chemical Abstracts Service, "Chemical Titles", No. 1, 5 Jan 1961, Columbus, Ohio. (issued semi-monthly.)

111. Chemical-Biological Coordination Center, "The Chemical-Biological Coordination Center of the National Academy of Sciences-National Research Council", National Academy of Sciences-National Research Council, Washington, D.C. Sep 1954, 33 p.

112. Cherry, C. [ed]. "Information Theory, Third London Symposium", Academic Press, New York, 1956, 401 p.

113. Cherry, C. [ed]. "Information Theory, Fourth London Symposium", (papers read at a symposium on information theory held at the Royal Institution, London, 29 Aug to 2 Sep 1960), Butterworths, London, 1961, 476 p.

114. Cheydleur, B.F. "Information Retrieval 1966", Datamation 7, 21-25 (1961).

115. Cheydleur, B.F. "SHIEF: a Realizable Form of Associative Memory", Amer. Documentation 14, 56-57 (1963).

116. Chonez, A. "Mecanisation Partielle des Taches Bibliographiques (1)", Rept. DOC-CEN/S-AFD-17, Centre d'Etudes Nucléaires de Saclay, Gif-sur-Yvette, France (June 1960) 10 p.

117. Chonez, A. "Mecanisation Partielle des Taches Bibliographiques (3)", Rept. DOC-CEN7S-AFD-19, Centre d'Etudes Nucléaires de Saclay, Gif-sur-Yvette, France (July 1960) 20 p.

118. Chonez, A. "Mecanisation Partielle des Taches Bibliographiques (6)", Rept. DOC-CEN/S-AFD-28, Centre d'Etudes Nucléaires de Saclay, Gif-sur-Yvette, France (Dec 1960) 15 p.

119. Chonez, N., A. Chonez and J. Iung, "Physindex: An Auto-Indexed Current List of Physics Literature Produced on IBM 1401 Computer", in H.P. Luhn [ed]. "Automation and Scientific Communication, Short Papers, Pt. 1", 1963, p. 31-32.

120. Citron, J., L. Hart and H. Ohlman, "A Permutation Index to the 'Preprints of the International Conference on Scientific Information'", SP-44, System Development Corp., Santa Monica, Cal. 1958, 140 p.

121. Citron, J., L. Hart and H. Ohlman, "A Permutation Index to the 'Preprints of the International Conference on Scientific Information'", Rept. No. SP-44, (Revised edition), System Development Corp., Santa Monica, Cal. 15 Dec 1959, 37 p.

122. Clapp, V. W. "Research in Problems of Scientific Information--Retrospect and Prospect", Amer. Documentation 14, 1-9 (1963).

123. Clark, L. L. "Some Computer Techniques in the Behavioral Sciences", in A. Kent [ed]. "Information Retrieval and Machine Translation, Pt. I", 1960, p. 445-446.

124. Cleverdon, C. W. "The ASLIB Cranfield Research Project on the Comparative Efficiency of Indexing Systems", ASLIB Proc. 12, 412-431 (1960).

125. Cleverdon, C. W. "Automation in Indexing", ASLIB Proc. 13, 107-109 (1961).

126. Cleverdon, C. W. "The Evaluation of Systems Used in Information Retrieval", in "Proceedings of the International Conference on Scientific Information", 1959, Vol I, p. 687-698.

127. Cleverdon, C. W. "Interim Report on the Test Programme of an Investigation into the Comparative Efficiency of Indexing Systems", ASLIB Cranfield Research Project, The College of Aeronautics, Cranfield, England, Nov 1960, 79 p.

128. Cleverdon, C. W. "An Investigation into the Comparative Efficiency of Information Retrieval Systems", UNESCO Bull. for Libraries 12, 267-270 (1958).

129. Cleverdon, C. W. "Report on Testing and Analysis of an Investigation into the Comparative Efficiency of Indexing Systems", ASLIB Cranfield Research Project, The College of Aeronautics, Cranfield, England, Oct 1962, 305 p.

130. Cleverdon, C. W., F. W. Lancaster and J. Mills, "Uncovering Some Facts of Life in Information Retrieval", Spec. Libraries 55, 84-91 (1964).

131. Cleverdon, C. W. and J. Mills, "The Analysis of Index Language Devices", presented at the ADI 1963 Annual Convention, 19 p.

132. Cleverdon, C. W. and J. Mills, "The Testing of Indexing Language Devices", College of Aeronautics, Cranfield, England, undated, 24 p. Also in ASLIB Proc. 15, 106-130 (1963).

133. Climenson, W. D., N. H. Hardwick and S. N. Jacobson, "Automatic Syntax Analysis in Machine Indexing and Abstracting", in "Machine Indexing", American U., 1962, p. 305-325. Also in Amer. Documentation 12, 178-183 (1961).

134. Coates, E. J. "Monitoring Current Technical Information with the British Technology Index", ASLIB Proc. 14, 426-437 (1962).

135. Committee on Scientific Information, Federal Council for Science and Technology, "Status Report on Scientific and Technical Information in the Federal Government", Washington, D. C. 18 June 1963, 18 p.

136. Connolly, T. F. "Author Participation In Indexing-From Primary Publication to Information Center", in H. P. Luhn [ed]. "Automation and Scientific Communication, Short Papers, Pt. 1", 1963, p. 35-36.

137. Conrad, G. M. "New Developments in the Merchandising of Biological Research Information", Amer. Scientist 50, 370A-378A (1962).

138. Conrad, G. M. and R. R. Gulick, "The Length and Structure of the Titles of Primary Biological Research Articles", Biological Abstracts, Philadelphia, Pa. 30 Sep 1962, 21 p.

139. Cook, C. M. "Automation Comes to the Bible", Christian Century 74, 892-894 (1957).

140. Cornelius, M. E. "Machine Input Problems for Machine Indexing: Alternatives and Practicalities", in "Machine Indexing", American U., 1962, p. 41-49.

141. Costello, J.C., Jr. "Storage and Retrieval of Chemical Research and Patent Information by Links and Roles in Du Pont", Amer. Documentation 12, 111-120 (1961).

142. Cox, G.J., C.F. Bailey and R.S. Casey, "Punch Cards for a Chemical Bibliography", Chem. and Eng. News 23, 1623-1626 (1945).

143. Coyaud, M. "Analyse Automatique de Documents Ecrits en Langue Naturelle vers un Language Documentaire (Le Syntol)", NATO Advanced Study Institute on Automatic Document Analysis, Venice, 7-20 July 1963. Preprint July 1963, 16 p.

144. Crane, E.J. and C.L. Bernier, "Indexing and Index Searching", in R.W. Casey, et al, "Punched Cards: Their Applications to Science and Industry", 1958, p. 510-527.

145. Crane, E.J. and C.L. Bernier, "An Overall Concept of Scientific Documentation Systems and Their Design", in "Proceedings of the International Conference on Scientific Information", 1959, Vol II, p. 1047-1069.

146. Crestadoro, A. "The Art of Making Catalogues of Libraries; A Method to Obtain in a Short Time a Most Perfect, Complete, and Satisfactory Catalogue of the British Museum Library, By a Reader Therein", The Literary, Scientific and Artistic Reference Office, London, 1856.

147. Dale, A.G. and N. Dale, "Some Clumping Experiments for Information Retrieval", Rept. no. LRC-64-WPIA, Linguistics Research Center, University of Texas, Austin, Tex. 1964, 11 p.

148. Damerau, F.J. "An Experiment In Automatic Indexing", IBM Research Rept., International Business Machines Corp., New York, 19 Feb 1963.

149. Danton, E.M. [ed]. "The Library of Tomorrow: A Symposium", The American Library Association, Chicago, 1939, 192 p.

150. Davis, D.D. "The Use of Punched-Tape Typewriters and Computers in the Centralized Information Processing at the USAEC Division of Technical Information Extension", in H.P. Luhn [ed]. "Automation and Scientific Communication, Short Papers, Pt. 2", 1963, p. 237-238.

151. Day, S. and I. Lebow, "New Indexing Pattern for Nuclear Science Abstracts", Amer. Documentation 11, 120-127 (1960).

152. de Grolier, E. "A Study of General Categories Applicable to Classification and Coding in Documentation", UNESCO, Paris, 1962, 248 p.

153. Dewey, H. "Punched Card Catalogs--Theory and Technique", Amer. Documentation 10, 36-50 (1959).

154. "Diabetes-Related Literature Index by Authors and By Keywords in the Title For the Year 1960", Diabetes 12, Supplement 1 (1963).

--- "Documentation, Indexing and Retrieval of Scientific Information", see U.S. Congress.

155. Documentation, Inc. "How to Ferret Out Information Electronically", Research Review, Office of Naval Research, Washington, D.C. 1956.

156. Documentation, Inc. "The Logic and Mechanics of Storage and Retrieval Systems", Technical rept. no. 14, Washington, D.C. Feb 1956, 37 p.

157. Documentation, Inc. "The Preparation of Manual Dictionaries of Association", Technical rept. no. 5, Washington, D.C. Apr 1954, 11 p.

158. Douglas Aircraft Company, Douglas Missiles and Space Library, "KWOC (Keyword-Out-Of-Context)", Santa Monica, Cal.

159. Dowell, N.G. and J.W. Marshall, "Experience With Computer-Produced Indexes", ASLIB Proceedings 14, 323-332 (1962).

160. Doyle, L.B. "Association Characteristics of Words in Text", Comm. Assoc. Computing Machinery 5, 223 (1962).

161. Doyle, L.B. "Discussion of a Proposed Study of Association Derived From Text", Rept. FN-6081, System Development Corp., Santa Monica, Cal. Dec 1961, 11 p.

162. Doyle, L.B. "Expanding the Editing Function in Language Data Processing", presented at ACM 1963 National Conference. Also Rept. no. SP-1268, System Development Corp., Santa Monica, Cal. 10 July 1963, 15 p.

163. Doyle, L.B. "Indexing and Abstracting By Association", Amer. Documentation 13, 378-390 (1962).

164. Doyle, L.B. "Is Relevance an Adequate Criterion in Retrieval System Evaluation?", Rept. no. SP-1262, System Development Corp., Santa Monica, Cal. July 1963, 6 p.

165. Doyle, L.B. "Library Science In the Computer Age", Rept. no. SP-141, System Development Corp., Santa Monica, Cal. 17 Dec 1959, 22 p.

166. Doyle, L.B. "A Method for Improving Organization In Large Computer-Generated Indexes", Rept. no. TM-628, System Development Corp., Santa Monica, Cal. 27 June 1961, 19 p.

167. Doyle, L.B. "The Microstatistics of Text", Rept. no. SP-1083, System Development Corp., Santa Monica, Cal. 21 Feb 1963, 36 p.

168. Doyle, L.B. "Programmed Interpretation of Text as a Basis for Information-Retrieval Systems", in "Proceedings of the Western Joint Computer Conference 1959", 1959, p. 60-63.

169. Doyle, L.B. "Semantic Road Maps For Literature Searchers", Rept. no. SP-199, System Development Corp., Santa Monica, Cal. 23 Jan 1961, 29 p. Also in J. Assoc. Computing Machinery 8, 553-578 (1961).

170. Doyle, L.B. "Statistical Analysis of Text in the Distant Future", Rept. no. SP-800, System Development Corp., Santa Monica, Cal. 30 Apr 1962, 20 p.

171. Doyle, L.B. "Statistical Semantics", in C. Popplewell [ed]. "Information Processing 1962", 1963, p. 335-336.

172. Dubester, H.J. "Mechanization of Subject Headings", Library Resources and Tech. Services 6, 230-234 (1962).

173. Durkin, R.E. and H.S. White, "Simultaneous Preparation of Library Catalogs for Manual and Machine Applications", Spec. Libraries 52, 231-237 (1961).

174. Dyson, G.M. and M.F. Lynch, "Chemical-Biological Activities, A Computer-Produced Express Digest", J. Chem. Documentation 3, 81-85 (1963).

175. Edmundson, H.P. "An Experiment in Abstracting Russian Text By Digital Computer", in H.P. Luhn [ed]. "Automation and Scientific Communication, Short Papers, Pt. 1", 1963, p. 83-84, 351 (Pt. 2).

176. Edmundson, H.P. "Linguistic Analysis in Machine-Translation Research", in M. Boaz [ed]. "Modern Trends in Documentation", 1959, p. 31-37.

177. Edmundson, H. P. "New Methods in Automatic Abstracting", Abstract, 1963 ACM National Conference, Denver, Colo. Aug 1963.

178. Edmundson, H. P. "Problems in Automatic Abstracting", in "Joint Man-Computer Indexing and Abstracting", Mitre SS-13, 1962, p. 1-15. Also in Comm. Assoc. Computing Machinery 7, 259-263 (1964).

179. Edmundson, H. P. [ed]. "Proceedings of the National Symposium on Machine Translation", Prentice-Hall, Englewood Cliffs, N.J. 1961, 525 p.

180. Edmundson, H. P., V.A. Oswald, Jr., and R.E. Wyllys, "Automatic Indexing and Abstracting of the Contents of Documents", Final rept. no. PRC-R-126, Planning Research Corp., Los Angeles, Cal. 31 Oct 1959, 133 p.

181. Edmundson, H. P. and R.E. Wyllys, "Automatic Abstracting and Indexing-Survey and Recommendations", Comm. Assoc. Computing Machinery 4, 226-234 (1961).

182. Eldridge, W.B. and S.F. Dennis, "The Computer as a Tool For Legal Research", Law and Contemporaty Problems 28, 77-99 (1963).

183. Eldridge, W.B. and S.F. Dennis, "Report of Status of the Joint American Bar Foundation--IBM Study of Electronic Methods Applied to Legal Information Retrieval", American Bar Foundation, Chicago, 1 Aug 1962, 7 p.

184. Ellegård, A. "Estimating Vocabulary Size", Word 16, 219-244 (1960).

185. Ellegård, A. "A Statistical Method for Determining Authorship", Gothenburg Studies in English, 13, Gothenburg, Sweden, 1962.

186. Ellison, J.W. "Nelson's Complete Concordance of the Revised Standard Version Bible", Nelson, New York, 1957, 2158 p.

187. Fair, E.M. "Inventions and Books - What of the Future?", Library J. 61, 47-51 (1936).

188. Fairthorne, R.A. "The Patterns of Retrieval", Amer. Documentation 7, 65-70 (1956).

189. Fairthorne, R.A. "Some Clerical Operations and Languages" in C. Cherry [ed]. "Information Theory, Third London Symposium", 1956, p. 111-120.

190. Fairthorne, R.A. "Towards Information Retrieval", Butterworths, London, 1961, 211 p.

191. Fano, R.M. "Information Theory and the Retrieval of Recorded Information", in J. Shera et al [eds]. "Documentation in Action", 1956, p. 238-244. (Also preprint mimeo), 16 May 1956.

192. Farley, E. "A New Permuted Title Index in the Social Sciences and the Humanities", Spec. Libraries 54, 557-562 (1963).

193. Farradane, J. "The Challenge of Information Retrieval", J. Documentation 17, 233-244 (1961).

194. Fasana, P.J. "Automating Cataloging Functions in Conventional Libraries", Lib. Resources & Tech. Services 7, 350-365 (1963).

195. Fasana, P.J. "Bibliographic Encoding: A Machine-Interpretable Format for Highly Structured Data", in H.P. Luhn [ed]. "Automation and Scientific Communication, Short Papers, Pt. 2", 1963, p. 325-326.

196. Fels, E.M. and J. Jacobs, "Linguistic Statistics of Indexing", Univ. Pittsburgh Law Review 24, 771-791 (1963).

--- First Congress on the Information System Sciences, see "Joint Man-Computer Indexing and Abstracting", and "Joint Man-Computer Languages", Mitre Corp. 1962.

197. Fishenden, R. M. "Methods by Which Research Workers Find Information", in "Proceedings of the International Conference on Scientific Information", Vol I, 1959, p. 163-179.

198. Ford, J. D., Jr. "Automated Content Analysis", Rept. no. TM-904, System Development Corp., Santa Monica, Cal. 19 Feb 1963, 12 p.

199. Foskett, D. J. "Two Notes on Indexing Techniques", J. Documentation 18, 188-192 (1962).

200. "Freeing the Mind", (articles and letters from the Times Literary Supplement During March-June, 1962), The Times Publishing Company, Ltd., London, 1962.

201. Freeman, R. R. "Automatic Retrieval and Selective Dissemination of References from Chemical Titles: Improving the Selection Process", in H. P. Luhn [ed]. "Automation and Scientific Communication, Short Papers, Pt. 2", 1963, p. 213-214.

202. Freeman, R. R. and G. M. Dyson, "Development and Production of Chemical Titles, A Current Awareness Index Publication Prepared with the Aid of a Computer", J. Chem. Documentation 3, 16-20 (1963).

203. Friedman, H. J. and G. M. Dyson, "Study of Semantics in Relation to the Machine Languages of Concepts", Chemical Abstracts Service, Columbus, Ohio, 1961, 57 p.

204. Friis, Th. "The Use of Citation Analysis as a Research Technique and its Implications for Libraries", South African Libraries 23, 12-15 (1955).

205. Gallagher, T. A. and P. J. Toomey, "A Case History in Automated Information Storage and Retrieval", in "Proceedings, Symposium on Materials Information Retrieval", 1963, p. 43-66.

206. Gardin, J. C. "Conferences de J. C. Gardin", preprint resumés of papers presented at the NATO Advanced Study Institute on Automatic Document Analysis, Venice, Italy, 7-20 July 1963, 24 p.

207. Gardin, J. C. "La Notion de Language Documentaire: Defense et Illustration", in "Conferences de J. C. Gardin", 1963, p. 12-15.

208. Gardin, J. C. "Pour une Classification Finie des Problemes de l'Automatique Documentaire", in "Conferences de J. C. Gardin", 1963, p. 1-5.

209. Gardin, J. C. "Stratégies Comparées en Matiere d'Analyse Documentaire Automatique", in "Conferences de J. C. Gardin", 1963, p. 6-11.

210. Garfield, E. "Association-of-Ideas Techniques in Documentation - Shepardizing the Literature of Science", Smith, Kline and French Laboratories, Philadelphia, Pa. Oct 1954, 11 p.

211. Garfield, E. "Breaking the Subject Index Barrier--A Citation Index For Chemical Patents", J. Pat. Office Soc. 39, 583-595 (1957).

212. Garfield, E. "Citation Indexes--New Paths to Scientific Knowledge", Chem. Bull. 43, 11-12 (1956).

213. Garfield, E. "Citation Indexes for Science: A New Dimension in Documentation Through Association of Ideas", Science 122, 108-111 (1955).

214. Garfield, E. "Citation Indexes in Sociological and Historical Research", Amer. Documentation 14, 289-291 (1963).

215. Garfield, E. "Generic Searching by Use of Rotated Formula Index", J. Chem. Documentation 3, 97-103 (1963).

216. Garfield, E. "Preliminary Report on the Mechanical Analysis of Information by Use of the 101 Statistical Punched Card Machines", preprint dated 19 Feb 1953. Also in Amer. Documentation 5, 7-12 (1954).

217. Garfield, E. "The Preparation of the Current List of Medical Literature by Punched-Card Methods", Welch Medical Library, Johns Hopkins Univ., Baltimore, Md. 1953.

218. Garfield, E. "Preparation of Printed Indexes by Automatic Punched-Card Equipment-A Manual of Procedures", Welch Medical Library, Johns Hopkins Univ., Baltimore, Md. 1953.

219. Garfield, E. "The Preparation of Printed Indexes by Automatic Punched-Card Techniques", Amer. Documentation 6, 68-76 (1955).

220. Garfield, E. "The Preparation of Subject Heading Lists by Punched Card Methods", J. Documentation 10, 1-10 (1954).

221. Garfield, E. "A Unified Index to Science", in "Proceedings of the International Conference on Scientific Information", 1959, Vol I, p. 461-474.

222. Garfield, E. and I.H. Sher, "Genetics Citation Index-Experimental Citation Indexes to Genetics with Special Emphasis on Human Genetics". Prepared by the Institute for Scientific Information, Eugene Garfield, director, Irving H. Sher, project director, Philadelphia, Pa. 1963, 864 p.

223. Garfield, E. and I.H. Sher, "New Factors in the Evaluation of Scientific Literature Through Citation Indexing", Amer. Documentation 14, 195-201 (1963).

224. Garvin, L. "Some Linguistic Aspects of Information Retrieval", in "Machine Indexing", American U., 1962, p. 134-143.

225. Gates, M.L., et al, "Punched Cards for Library Records", Library J. 71, 1783-1786 (1946).

226. Giallanza, F.V. and J.H. Kennedy, "Key-Word-in-Title (KWIT) Index for Reports", Rept. no. UCRL-6782, Lawrence Radation Laboratory, Univ. of California, Livermore, Cal. 14 May 1962, 8 p.

227. Giuliano, V.E. "Analog Networks for Word Association", IRE Trans. Military Electronics MIL-7, 221-234 (1963).

228. Giuliano, V.E. "Automatic Message Retrieval by Associative Techniques", in "Joint Man-Computer Languages", Mitre SS-10, 1962, 1-44 p.

229. Giuliano, V.E. and P.E. Jones, "Linear Associative Information Retrieval", Rept. no. CACL-2, Arthur D. Little, Inc., Cambridge, Mass., Nov 1962, 240 p. Also in P.W. Howerton and D.C. Weeks [eds]. "Vistas in Information Handling", 1963, p. 30-54.

230. Giuliano, V.E., et al. "Automatic Message Retrieval", Studies for the Design of an English Command and Control Language System (final report) EST-TDR-63-673, Arthur D. Little, Inc., Cambridge, Mass. Nov 1963, 187 p.

231. Giuliano, V.E., et al. "Studies for the Design of an English Command and Control Language System", Rept. no. ESD-TR-62-45, Arthur D. Little, Inc., Cambridge, Mass. June 1962, 118 p.

232. Glass, B. and S.H. Norwood, "How Scientists Actually Learn of Work Important to Them", in "Proceedings of the International Conference on Scientific Information", 1959, Vol I, p. 195-197.

233. Goldwyn, A. J. "The Place of Indexing in the Design of Information Systems Tests", in H. P. Luhn [ed]. "Automation and Scientific Communication, Short Papers, Pt. 2", 1963, p. 321-322.

234. Good, I. J. "Speculations Concerning Information Retrieval", Rept. no. RC-78, IBM Research Center, Yorktown Heights, N. Y. 10 Dec 1958, 14 p.

235. Goodman, F. L. "A Citation Index for Literature on New Educational Media", in H. P. Luhn [ed]. "Automation and Scientific Communication, Short Papers, Pt. 1", 1963, p. 33-34.

236. Gordon, L. and R. Slowinski, "The Evolution of Medical Terminology Through Electronic Equipment and Photographic Reproduction", in H. P. Luhn [ed]. "Automation and Scientific Communication, Short Papers, Pt. 1", 1963, p. 55.

237. Gottschalk, L.A. [ed]. "Comparative Psycholinguistic Analysis of Two Psychotherapeutic Interviews", International Universities Press, New York, 1961, 221 p.

238. Green, B.F., Jr., A.K. Wolf, C. Chomsky and K. Laughery, "Baseball: An Automatic Question-Answerer", in "Proceedings of the Western Joint Computer Conference", 1961, Vol 19, p. 219-224.

239. Greer, F. L. "The User Approach to Information Systems", General Electric Co., Information Systems Operation, Washington, D. C. May 1963, 40 p.

240. Greer, F. L. "Word Usage and Implications for Storage and Retrieval", General Electric Co., Information Systems Operation, Washington, D. C. July 1962, 74 p.

241. Griffin, M. "Printed Book Catalogs", Rev. de la Doc. 28, 8-17 (1961).

242. Griffin, M. "Printed Book Catalogs", Spec. Libraries 51, 496-499 (1960).

243. Grimes, J. E. and M. Alvarez, "The S. I. L. Concordance Program", presented at the meeting of the Linguistic Society of America, Austin, Tex. July 1961.

244. Grosch, H.R.J. "The Nature of Information Retrieval", in M. Boaz [ed]. "Modern Trends in Documentation", 1959, p. 13-22.

245. Gull, C.D. "A Punched-Card Method for the Bibliography, Abstracting, and Indexing of Chemical Literature", J. Chem. Ed. 23, 500-507 (1946).

246. Gull, C.D. "Seven Years of Work on the Organization of Materials in the Special Library", Amer. Documentation 7, 320-329 (1956).

247. Gull, C.D. "A Summary of Applications of Punched Cards as They Affect Special Libraries", Spec. Libraries 38, 208-212 (1947).

248. Gurk, H.M. and J. Minker, "The Design and Simulation of an Information Processing System", J. Assoc. Computing Machinery 8, 260-270 (1961).

249. Halliday, M.A.K. "The Linguistic Basis of a Mechanical Thesaurus", Mech. Translation 3, 81-88 (1956).

250. Hammond, W. "Convertibility of Indexing Vocabularies", in "The Literature of Nuclear Science: Its Management and Use", U.S. Atomic Energy Commission, 1962, Section III-3, p. 223-234.

251. Hammond, W., S. Rosenborg and J. Jaster, "A Search Strategy for Retrieving Legal Information", Technical Rept. no. IR-2, Datatrol Corp., Silver Spring, Md. Dec 1962, 19 p.

252. Hammond, W. and S. Rosenborg, "Experimental Study of Convertibility Between Large Technical Indexing Vocabularies", Technical Rept. no. IR-1, Datatrol Corp., Silver Spring, Md. Aug 1962.

253. Hardkopf, J.C. "Cybernetics and The Library", Library J. 76, 999-1001 (1951).

254. Harris, Z.S. "Linguistic Transformations for Information Retrieval", in "Proceedings of the International Conference on Scientific Information, 1959, Vol 2, p. 937-950.

255. Hart, H.C. "Re: Citation System for Patent Office", J. Pat Office Society 31, 714 (1949).

256. Hart, L.D. and G.R. Bach, "Natural Language Indexing by Means of Data-Processing Machines", (Observation of the Growth of Perception Protocol), Rept. no. SP-78, System Development Corp., Santa Monica, Cal. June 1959, 19 p.

257. Hattery, L.H. and E.M. McCormick [eds]. "Information Retrieval Management", American Data Processing, Inc., Detroit, Mich. 1962, 151 p.

258. Hays, D.G. "Linguistic Research at the RAND Corporation", in H.P. Edmundson [ed]. "Proceedings of the National Symposium on Machine Translation", 1961, p. 13-25.

259. Heiliger, E. "Application of Advanced Data Processing Techniques to University Library Procedures", Spec. Libraries 53, 472-475 (1962).

260. Heller, W. "Applied Information Management System", in H.P. Luhn [ed]. "Automation and Scientific Communication, Short Papers, Pt. 2", 1963, p. 161-162.

261. Heller, E.W. "Applied Information Management System User's Manual", Rept. no. TM-1201/000/60, System Development Corp., Santa Monica, Cal. 23 Apr 1963, 26 p.

262. Helyar, L.E.J. "Summing Up and Conclusions", ASLIB Proc. 13, 110-111 (1961).

263. Henderson, M.M.B. "Organizations Active in Machine Indexing Research", in "Machine Indexing", American U., 1962, p. 22-39.

264. Herner, S. "Deep Indexing by Manual Permutation Methods", preprint for Annual Meeting, A.D.I., 1963, Herner and Co., Washington, D.C. 22 Aug 1963, 11 p.

265. Herner, S. "The Information-Gathering Habits of American Medical Scientists", in "Proceedings of the International Conference on Scientific Information", 1959, Vol I, p. 277-285.

266. Herner, S. "Methods of Organizing Information for Storage and Searching", Amer. Documentation 13, 3-14 (1962).

267. Herner, S. "The Role of Thesauri in the Convergence of Word and Concept Indexing", in H.P. Luhn [ed]. "Automation and Scientific Communication, Short Papers, Pt. 2", 1963, p. 183-184.

268. Hessel, A. "A History of Libraries", (translated, with supplementary material, by Reuben Preiss), Scarecrow Press, New Brunswick, N.J. 1955, 198 p.

269. Heumann, K.F. "The Big Black Box at Your Beck and Call", Spec. Libraries 51, 483-484 (1960).

270. Heumann, K.F. "The Chemical-Biological Coordination Center", National Academy of Sciences - National Research Council, News Report 2, 67-69 (1952).

271. Heumann, K.F. and E. Dale, "Statistical Survey of Chemical Structure", in G.L. Peakes, et al [eds]. "A Progress Report in Chemical Literature Retrieval", 1957, p. 201-214.

272. Hillman, D.J. "Mathematical Theories of Relevance with Respect to Systems of Automatic and Manual Indexing", in H.P. Luhn [ed]. "Automation and Scientific Communication, Short Papers, Pt. 2", 1963, p. 323-324.

273. Hines, T.C. "Machine Arrangement of Alphanumeric Concordance, Thesaurus, and Index Entries: The Need for Compatible Standard Rules", in H.P. Luhn [ed]. "Automation and Scientific Communication, Short Papers, Pt. 1", 1963, p. 7-8.

274. Hocken, S. "Disseminating Current Information", Spec. Libraries 53, 93-95 (1962).

275. Hoffman, W. [ed]. "Digital Information Processors, Selected Articles on Information Processing", Interscience Publishers, New York, 1962, 740 p.

276. Horty, J.F. "Electronic Data Retrieval of Law", Current Business Studies 36, 35-46 (1961).

277. Horty, J.F. "Experience with the Application of Electronic Data Processing Systems in General Law", M.U.L.L. (Modern Uses of Logic in Law) 60D, 158-168 (1960).

278. Horty, J.F. "The Keyword in Combination Approach", M.U.L.L. (Modern Uses of Logic in Law) 62M, 54 (1962).

279. Horty, J.F. "Searching Statutory Law by Computer, Interim Report No. 1 to Council on Library Resources, Inc." Health Law Center, Univ. of Pittsburgh, Pa. undated.

280. Horty, J.F. and T.B. Walsh, "Use of Flexowriters to Prepare Large Amounts of Alphabetic Legal Data for Computer Retrieval", in H.P. Luhn [ed]. "Automation and Scientific Communication, Short Papers, Pt. 2", 1963, p. 259-260.

281. "How to Use Shepard's Citations", Shepard's Citations, Inc., Colorado Springs, Colo. 1873 and subsequently.

282. Howerton, P.W. "The Application of Modern Lexicographic Techniques to Machine Indexing", in "Machine Indexing", American U., 1962, p. 326-330.

283. Howerton, P.W. and D.C. Weeks [eds]. "Vistas in Information Handling", Vol I, "The Augmentation of Man's Intellect by Machine", Spartan Books, Washington, D.C. 1963.

284. Hughes, C.J. "A Critical Comparison of Some Typical Data Retrieving Systems", in G. Salton [ed]. "Information Storage and Retrieval, no. ISR-2", 1 Sep 1962, p. IV-1 to IV-28.

285. "IBM Punched-Card Accounting Is Adapted to Make Scholarly Indexes", Publishers Weekly 170, 2150-2152 (1956).

286. "Information Processing", Proceedings of the International Conference on Information Processing, UNESCO, Paris, 13-15 June 1959, Butterworths, London, 1960.

287. International Business Machines Corp. "General Information Manual: Keyword-In-Context (KWIC) Indexing", White Plains, N.Y. 1962, 21 p.

288. International Business Machines Corp. "General Information Manual: Mechanized Library Procedures", White Plains, N.Y. undated, 19 p.

289. International Business Machines Corp. Advanced Systems Development Division, "ACSI-matic Auto-Abstracting Project", Final Report, Vol 1, Yorktown Heights, N.Y. 22 Feb 1960, 217 p.

290. International Business Machines Corp. Advanced Systems Development Division, "ACSI-matic Auto-Abstracting Project", Final Report, Vol 3, Yorktown Heights, N.Y. 31 Mar 1961, 126 p.

291. Iung, J. and N. Vandeputte, "Les Donnees Documentaires, Leur Manipulation. Etude Preliminairea l'Utilisation des Machines Mecanographiques et Logiques", (DOC-CEN/ S-AFD 22) Centre d' Etudes Nucléaires de Saclay, Gif sur Yvette, France, Aug 1960, 27 p.

292. Jacobson, S.N. "Paragraph Analysis: Novel Technique for Retrieval of Portions of Documents", in H.P. Luhn [ed]. "Automation and Scientific Communication, Short Papers, Pt. 2", 1963, p. 191-192.

293. Jacoby, J. and V. Slamecka, "Indexer Consistency Under Minimal Conditions", Documentation, Inc., Bethesda, Md. Nov 1962, lv.

294. Jaffe, J. "Computer Analysis of Verbal Behavior in Psychiatric Interviews", presented at the annual meeting of the Association for Research in Nervous and Mental Disease, New York, Dec 1962, Columbia University, College of Physicians and Surgeons, New York, undated, 23 p.

295. Jaffe, J. "Dyadic Analysis of Two Psychotherapeutic Interviews", in L.A. Gottschalk [ed]. "Comparative Psycholinguistic Analysis of Two Psychotherapeutic Interviews", 1961.

296. Jaffe, J. "Electronic Computers in Psychoanalytic Research", in J.H. Masserman [ed]. "Science and Psychoanalysis", Vol VI, 1958.

297. Jaffe, J. "Electronic Computers in Psychoanalytic Research", presented at the Annual Meeting of the Academy of Psychoanalysis, 4-6 May 1952, Toronto, Canada, undated, 20 p.

298. Jahoda, G. "The Development of a Combination Manual and Machine-Based Index to Research and Engineering Reports", Spec. Libraries 53, 74-78 (1962).

299. Janaske, P.C. "Manual Preparation of a Permuted-Title Index", BSCP Communique, 7-62, Biological Abstracts, Philadelphia, Pa. June 1962, 15 p.

300. Johnson, H.T. "A Polydimensional Scheme for Information Retrieval", Amer. Documentation 13, 90-92 (1962).

301. Johnson, H.T. "A Program for Dissemination of Specific Data on Materials", in H.P. Luhn [ed]. "Automation and Scientific Communication, Short Papers, Pt. 2", 1963, p. 295-296.

302. "Joint Man-Computer Indexing and Abstracting", First Congress on the Information System Sciences, 1st draft-Information System Science and Engineering, Mitre SS-13, Mitre Corp., Bedford, Mass. 1962, 73 p.

303. "Joint Man-Computer Languages", Proceedings of the First Congress on the Information System and Engineering, Mitre SS-10, Mitre Corp., Bedford, Mass. 1962, 105 p.

304. Jones, P.E. "Research on a Linear Network Model and Analog Device for Associative Retrieval", in H.P. Luhn [ed]. "Automation and Scientific Communication, Short Papers, Pt. 2", 1963, p. 211-212.

305. Joyce, T. and R.M. Needham, "The Thesaurus Approach to Information Retrieval", Amer. Documentation 9, 192-197 (1958).

306. Juncosa, M. L. "Symposium on Optimum Routing in Large Networks", in C. Popplewell [ed]. "Information Processing 1962", 1963, p. 716-721.

307. Kansas University Libraries, "Kansas Slavic Index, Current Titles, Social Sciences, Humanities, Permuted Title Index, Computer-based, 1963", Lawrence, Kans. 1963, 153 p.

308. Katter, R. V. "Language Structure and Interpersonal Commonality", System Development Corp., Rept. no. SP1185/000/01, Santa Monica, Cal. 17 June 1963, 30 p.

309. Kehl, W. B., J. F. Horty, C. R. I. Bacon and D. S. Mitchell, "An Information Retrieval Language for Legal Studies", Comm. Assoc. Computing Machinery 4, 380-389 (1961).

310. Kennedy, R. A. "Library Applications of Permutation Indexing", J. Chem. Documentation 2, 181-185 (1962).

311. Kennedy, R. A. "Mechanized Title Word Indexing of Internal Reports", in "Machine Indexing", American U., 1962, p. 112-132.

312. Kennedy, R. A. "Writing Informative Titles for Technical Papers--A Guide to Authors", in H. P. Luhn [ed]. "Automation and Scientific Communication, Short Papers, Pt. 2", 1963, p. 133-134.

313. Kent, A. [ed]. "Information Retrieval and Machine Translation (Part I)", Interscience Publishers, Inc., New York, 1960, 685 p.

314. Kent, A. "Information Retrieval-Review and Prospectus", in C. Popplewell [ed]. "Information Processing 1962", 1963, p. 267-272.

315. Kent, A. "Textbook on Mechanized Information Retrieval", Interscience Publishers, Inc., New York, 1962, 268 p.

316. Keppel, F. P. "Looking Forward, a Fantasy", in Danton, E. M. [ed]. "The Library of Tomorrow", 1939, p. 1-11.

317. Kessler, M. M. "Analysis of Bibliographic Sources in a Group of Physics-Related Journals", Rept. no. R-4, M. I. T., Lincoln Laboratory, Lexington, Mass. 6 Aug 1962.

318. Kessler, M. M. "Bibliographic Coupling Between Scientific Papers", Rept. no. R-2, M. I. T. Lincoln Laboratory, Lexington, Mass. 9 July 1962, 29 p. Also in Amer. Documentation 14, 10-25 (1963).

319. Kessler, M. M. "Bibliographic Coupling Extended in Time: Ten Case Histories", Rept. no. R-5, M. I. T. Lincoln Laboratory, Lexington, Mass. 20 Aug 1962, 37 p.

320. Kessler, M. M. "Comparison of the Results of Bibliographic Coupling and Analytic Subject Indexing", Rept. no. R-7, M. I. T. Lincoln Laboratory, Lexington, Mass. 28 Jan 1963, 30 p.

321. Kessler, M. M. "An Experimental Study of Bibliographic Coupling Between Technical Papers", Rept. no. R-1, M. I. T. Lincoln Laboratory, Lexington, Mass. 21 Nov 1961 (rev. 15 June 1962), 13 p.

322. Kessler, M. M. "Technical Information Flow Patterns", in "Proceedings of the Western Joint Computer Conference 1961", Vol 19, p. 247-257.

323. Kessler, M. M. and F. E. Heart, "Concerning the Probability That a Given Paper Will be Cited", Rept. no. R-6, M. I. T. Lincoln Laboratory, Lexington, Mass. 5 Nov 1962, 19 p.

324. Kilgour, F. G., R. T. Esterquest and T. P. Fleming, "Computerization of Book Catalogues at the Columbia, Harvard and Yale Medical Libraries", in H. P. Luhn [ed]. "Automation and Scientific Communication, Short Papers, Pt. 2", 1963, p. 299-300.

325. Klein, S. and R.F. Simmons, "Automated Analysis and Coding of English Grammar for Information Processing Systems", SDC Doc. SP-490, System Development Corp., Santa Monica, Cal. Sep 1961.

326. Kochen, M. "Adaptive Mechanics in Digital Concept-Processing", in Kochen, et al, "Adaptive Man-Machine Concept-Processing", 1962, App. V.

327. Kochen, M. "Techniques for Document Retrieval Research: State of the Art", IBM Research report RC 947, International Business Machines Corp., Yorktown Heights, N.Y. 31 Dec 1963, 24 p.

328. Kochen, M., C.T. Abraham and E. Wong, "Adaptive Man-Machine Concept-Processing", Final report, Thomas J. Watson Research Center, IBM, Yorktown Heights, N.Y. (AFCRL 62-397) 14 June 1962, 156 p.

329. Kochen, M., C.T. Abraham, E. Wong and H. Bohnert, "High-Speed Document Perusal", Final Technical Report, 1 Apr 1961 - 1 Apr 1962, Thomas J. Watson Research Center, IBM, Yorktown Heights, N.Y. 1 May 1962, lv.

330. Koelewijn, G.J. "Recent Developments in Western Europe in the Field of the Automation of Document Retrieval Systems", Rev. Int. Doc. 29, 42-47 (1962).

331. Korotkin, A.L. and L.H. Oliver, "The Effect of Subject Matter Familiarity and the Use of an Indexing Aid Upon Inter-Indexer Consistency", General Electric Company, Information Systems Operation, Bethesda, Md. 14 Feb 1964, 17 p.

332. Korotkin, A.L. and L.H. Oliver, "A Method for Computing Indexer Consistency", General Electric Company, Information Systems Operation, Bethesda, Md. 14 Feb 1964, 8 p.

333. Kraft, D.H. "Comparison of Keyword-in-Context (KWIC) Indexing of Titles With a Subject Heading Classification System", presented at the Annual Meeting of the American Documentation Institute, Hollywood by the Sea, Fla. 11-14 Dec 1962. International Business Machines Corp., Chicago, 1962.

334. Kraft, D.H. "An Operational Selective Dissemination of Information (SDI) System for Technical and Non-Technical Personnel Using Automatic Indexing Techniques", in H. P. Luhn [ed]. "Automation and Scientific Communication, Short Papers, Pt. 1", 1963, p. 69-70.

335. Kuhns, J.L. "An Application of Logical Probability to Problems in Automatic Abstracting and Information Retrieval", in "Joint Man-Computer Indexing and Abstracting", Mitre SS-13, 1962, p. 17-36.

336. Kuhns, J.L. "Mathematical Analysis of Correlation Clusters", in Ramo-Wooldridge, "Word Correlation and Automatic Indexing", Prog. rept. no. 2, 1959, Appendix D.

337. Kuipers, J.W. "Summary of Project Activities", (1 Nov 1958 - 31 Jan 1961), Contract NSF-C88, Itek Doc. IL-4000-17, Itek Corp., Lexington, Mass. 28 Feb 1961, 46 p.

338. Kuipers, J.W. and T.M. Williams, "A Program of Research and Development on Information Searching Systems", Chapter VI, Itek Doc. IL-4000-18, Itek Corp., Lexington, Mass. Aug 1958.

339. Kuno, S. and A.G. Oettinger, "Multiple-Path Syntactic Analyzer", in C.M. Popplewell [ed]. "Information Processing 1962", 1963, p. 306-312.

340. Kuno, S. and A.G. Oettinger, "Prospects for Automatic Processing of English Language Data", in H.P. Luhn [ed]. "Automation and Scientific Communication, Short Papers, Pt. 1", 1963, p. 5-6.

341. Kuno, S. and A.G. Oettiner, "Syntactic Structure and Ambiguity of English", in "Proceedings of the Fall Joint Computer Conference", 1963, p. 397-418.

342. Kyle, B. "Consistency Analysis of Two Indexers in Using K. C. for Political Science Material", (mimeo.), National Book League, London, Mar 1962, 6 p.

343. Lalley, J.M. "A Treasure Lost in Unread Wedges", The Washington Post, 2 Dec 1962, p. A10.

344. Lancaster, F.W. and J. Mills, "Testing Indexes and Index Language Devices: ASLIB Cranfield Project", Amer. Documentation 15, 4-13 (1964).

345. Lane, B.B. "Key Words In--and Out of --Context", Spec. Libraries 55, 45-46 (1964).

346. Langevin, R.A. and M. Owens, "Application of Automatic Syntactic Analysis to the Nuclear Test Ban Treaty", Rept. no. TO-B 63-71, Technical Operations Research, Burlington, Mass. 16 Aug 1963, 26 p.

347. Langleben, M.M. and A.L. Shumilina, "On Translation of Titles of Chemical Papers Into Information Languages", JPRS 13173, Joint Publication Research Service, No. 80, Washington, D.C. 27 Mar 1962, p. 99-119.

348. Larkey, S.V. "The Army Medical Library Research Project at the Welch Medical Library", Bull. Med. Lib. Assoc. 37, 121-124 (1949).

349. Larkey, S.V. "Cooperative Information Processing-Prospectus, Medicine", in J.H. Shera, et al. "Documentation in Action", 1956, p. 301-306.

350. Larkey, S.V. "Report on the Research Project of the Welch Medical Library", Bull. Med. Lib. Assoc. 39, 87-89 (1951).

351. Larkey, S.V. "The Welch Medical Library Indexing Project", Bull. Med. Lib. Assoc. 41, 32-40 (1953).

352. Ledley, R.S. "Tabledex: A New Coordinate Indexing Method for Bound Book Form Bibliographies", in "Proceedings of the International Conference on Scientific Information", 1959, Vol II, p. 1221-1243.

353. Lefkovitz, D. "Automatic Stratification of Descriptors", Moore School Rept. no. 64-03, University of Pennsylvania, Philadelphia, Pa. 15 Sep 1963, lv.

354. Lemmon, A. "Report on a Syntactic Analysis Program for Information Retrieval", Sect. II in G. Salton, "Information Storage and Retrieval, Scientific rept. no. ISR-2", Sep 1962, p. II-1 to II-27.

355. LeRoy, A. and P. Braffort, "Notice Relative a l'Elaboration D'un Codage par Phrases-Clés pour la Programmation d'un Systeme de Selection Automatique des Documents", Note C.E.A. no. 278, Centre d'Etudes Nucléaires de Saclay, Gif-sur-Yvette, France, May 1959, 20 p.

356. Lesk, M. "Attempts to Cluster Documents with Citation Data", Section VI, in G. Salton, "Information Storage and Retrieval", rept. no. ISR-3, 1 Apr 1963, p. VI-1 to VI-6.

357. Lesk, M. "A Comparison of Citation Data for Open and Closed Document Collections", Section V, in G. Salton, "Information Storage and Retrieval", rept. no. ISR-3, 1 Apr 1963, p. V-1 to V-10.

358. Lesk, M. and E. Storm, "A Computer Experiment for Sentence Extraction", Section I, in G. Salton, "Information Storage and Retrieval", rept. no. ISR-2, 1 Sep 1962, p. I-1 to I-34.

359. Levery, F. "An Experiment in Automatic Indexing of French Language Documents", in H.P. Luhn [ed]. "Automation and Scientific Communication, Short Papers, Pt. 1", 1963, p. 235-236.

360. Lilley, O. L. "Evaluation of the Subject Catalog: Criticisms and a Proposal", Amer. Documentation 5, 41-60 (1954).

361. Linder, L. H. "Indexing Costs for 10, 000 Documents", in H. P. Luhn [ed]. "Automation and Scientific Communication, Short Papers, Pt. 2", 1963, p. 147-148.

362. Linder, L. H. "Permutation Indexing as an Interim Means of Information Control", in "Proceedings of the March AFBMD Conference", 1960, p. 99-102.

363. Lindsay, R. K. "The Reading Machine Problem", unpublished Ph. D. dissertation, Graduate School of Industrial Administration, Carnegie Institute of Technology, Pittsburgh, Pa. 1960, 89 p.

364. Lipetz, B. A. "Compilation of an Experimental Citation Index from Scientific Literature", Technical rept. no. IL4000-19, Itek Corp., Lexington, Mass. June 1961, lv. Also in Amer. Documentation 13, 251-266 (1962).

365. Lipetz, B. A. "Compilation of an Experimental Citation Index from Scientific Literature with the Aid of Punched Card Equipment", Itek rept. no. RPIS-60-13, Itek Corp., Lexington, Mass. 1 Oct 1960, 32 p.

366. Lipetz, B. A. "Design of an Experiment for Evaluation of the Citation Index as a Reference Aid", in H. P. Luhn [ed]. "Automation and Scientific Communication, Short Papers, Pt. 2", 1963, p. 265-266.

367. Lipetz, B. A. "A Successful Application of Punched Cards in Subject Indexing", Amer. Documentation 11, 241-246 (1960).

368. Lipetz, B. A., D. E. Sparks and P. J. Fasana, "Techniques for Machine-Assisted Cataloging of Books", Report IL-9028-08, Spec. rept. no. 3 on Contract AF 19(604) 8438, Itek Corp., Lexington, Mass. 1962, 65 p.

--- "The Literature of Nuclear Science". See U. S. Atomic Energy Commission.

369. Lockheed Aircraft Corp. "An Evaluation of Information Retrieval Systems", Memo rept. no. 7170, Burbank, Cal. 30 Sep 1959, 114 p.

370. Loftus, H. E. "Automation in the Library - an Annotated Bibliography", Amer. Documentation 1, 110-126 (1956).

371. Luhn, H. P. "Auto-Encoding of Documents For Information Retrieval Systems", in M. Boaz [ed]. "Modern Trends in Documentation", 1959, p. 45-58.

372. Luhn, H. P. "Automated Intelligence Systems", in L. H. Hattery and E. M. McCormick [eds]. "Information Retrieval Management", 1962, p. 92-100.

373. Luhn, H. P. "Automated Intelligence Systems--Some Basic Problems and Prerequisites for Their Solution", in E. A. Tomeski, et al, "Clarification, Unification and Integration of Information Storage and Retrieval". 1961, p. 3-20.

374. Luhn, H. P. "The Automatic Creation of Literature Abstracts (Auto-Abstracts)", IBM J. Research and Development 2, 159-165 (1958). Also pub. in IRE National Convention Record, 1958, Institute of Radio Engineers, New York, 1958, Vol 6, Pt. 10, p. 20-24.

375. Luhn, H. P. "The Automatic Derivation of Information Retrieval Encodements from Machine-Readable Texts", International Business Machines Corp., Yorktown Heights, N. Y. 1959, 9 p. Also in A. Kent, "Information Retrieval and Machine Translation", Pt. II, 1961, p. 1021-1028.

376. Luhn, H. P. [ed]. "Automation and Scientific Communication, Short Papers, Pt. 1", American Documentation Institute, Washington, D. C. 1963, p. 1-128.

377. Luhn, H. P. [ed]. "Automation and Scientific Communication, Short Papers, Pt. 2", American Documentation Institute, Washington, D. C. 1963, p. 129-384.

378. Luhn, H. P. "A Business Intelligence System", IBM J. Research and Development 2, 314-319 (1958).

379. Luhn, H. P. "An Experiment in Auto-Abstracting: Auto-Abstracts of Area 5 Conference Papers International Conference on Scientific Information", IBM Research Center, Yorktown Heights, N. Y. 17 Nov 1958, 18 p.

380. Luhn, H. P. "General Rules for Creating Machinable Records for Libraries and Special Reference Files", Rept. no. 419, International Business Machines Corp., Yorktown Heights, N. Y. 30 Sep 1959.

381. Luhn, H. P. "Keyword-In-Context Index for Technical Literature (KWIC Index)", presented at American Chemical Society, Division of Chemical Literature at Atlantic City, N. J. 14 Sep 1959. Rept. no. RC 127, International Business Machines Corp., Yorktown Heights, N. Y. 1959, 16 p. Also in Amer. Documentation 11, 288-295 (1960).

382. Luhn, H. P. "Machinable Bibliographic Records as a Tool for Improving Communication of Scientific Information", paper presented at the 10th Pacific Scientific Congress, International Business Machines Corp., White Plains, N. Y. 1961.

383. Luhn, H. P. "A New Method of Recording and Searching Information", Amer. Documentation 4, 14-16 (1953).

384. Luhn, H. P. "Potentialities of Auto-Encoding of Scientific Literature", Res. rept. RC-101, International Business Machines Corp., Yorktown Heights, N. Y. 15 May 1959, 22 p.

385. Luhn, H. P. "A Statistical Approach to Mechanized Encoding and Searching of Literary Information", IBM J. Research and Development 1, 309-317 (1957).

386. Luhn, H. P. and P. James, "Bibliography and Index, Literature on Information Retrieval and Machine Translation, Titles Indexed by Keywords-In-Context System", The Service Bureau Corp., New York, Sep 1958, 42 p.

387. Lykoudis, P. S., P. E. Liley and Y. S. Touloukian, "Analytical Study of a Method for Literature Search in Abstracting Journals", in "Proceedings of the International Conference on Scientific Information", 1959, Vol I, p. 351-375.

388. Lyons, J. C. "A Search Strategy for Legal Retrieval", paper presented at American Bar Association Annual Meeting, San Francisco, Cal. 7 Aug 1962.

--- "Machine Indexing: Progress and Problems". See The American University.

389. MacMillan, J. T. and I. Welt, "A Study of Indexing Procedures in a Limited Area of the Medical Sciences", Amer. Documentation 12, 27-31 (1961).

390. MacQuarrie, C. "IBM Book Catalog", Library J. 82, 630-634 (1957).

391. MacWatt, J. A. "The Future and Three New Index Services", UNESCO Bull. Lib. 16, 187-190 (1962).

392. Maizell, R. E. "Value of Titles for Indexing Purposes", Rev. Doc. 27, 126-127 (1960).

393. Marckworth, M. L. "Dissertations in Physics, An Indexed Bibliography of All Doctoral Theses Accepted by American Universities, 1861-1959". Compiled with the assistance of the Staff at the Advanced Systems Development Division and Research Laboratories, IBM Corp., San Jose, Cal. Stanford University Press, 1961, 803 p.

394. Markus, J. V. "State of the Art of Published Indexes", Amer. Documentation 13, 15-30 (1962).

395. Maron, M. E. "Automatic Indexing: An Experimental Inquiry", Rept. no. P-2180, 1 Sep 1960 (rev. 2 Feb 1961), 31 p. Also in "Machine Indexing", American U., 1962, p. 236-265. Also in J. Assoc. Computing Machinery 8, 404-417 (1961).

396. Maron, M. E. "Probability and the Library Problem", Behavioral Science 8, 250-257 (1963).

397. Maron, M. E. and J. L. Kuhns, "On Relevance, Probabilistic Indexing and Information Retrieval", J. Assoc. for Computing Machinery 7, 216-244 (1960).

398. Maron, M. E., J. L. Kuhns and L. C. Ray, "Probabilistic Indexing. A Statistical Technique for Document Identification and Retrieval", Tech. memo no. 3, Thompson Ramo Wooldridge, Los Angeles, Cal. June 1959, 91 p.

399. Marthaler, M. P. "Current Research in Automatic Scientific Documentation", MHO/PA/231. 63, UNESCO working party in Scientific Documentation, No. 2: Automatic Documentation-Storage and Retrieval, Moscow, 11-16 Nov 1963. Preprint 11 Oct 1963, 69 p.

400. Martin, A. F. "IBM Catalog for the King County Public Library", Master's Thesis, Western Reserve Library School, Cleveland, Ohio, 1953, lv.

401. Massachusetts Institute of Technology Libraries, "KWIC Index to the Science Abstracts of China", prepared for the Symposium on Sciences of Communist China held by the American Association for the Advancement of Science, 26-27 Dec 1960 with the aid of a grant from the National Science Foundation. Cambridge, Mass. first edition Dec 1960, 134 p.

402. Masserman, J. H. [ed]. "Science and Psychoanalysis", Vol VI, Grone and Stratten, New York, 1958, lv.

403. Masterman, M. M. "The Potentialities of a Mechanical Thesaurus", presented at 2nd International Conference on Mechanical Translation, M. I. T., 16-20 Oct 1956. Also in Mech. Translation 3, 36 (1956).

404. Masterman, M. M. "The Thesaurus in Syntax and Semantics", Mech. Translation 4, 35-43 (1957).

405. Masterman, M. M., R. M. Needham and K. Spärck-Jones, "The Analogy Between Mechanical Translation and Information Retrieval", in "Proceedings of the International Conference on Scientific Information", 1959, Vol II, p. 917-955.

406. Mauchly, J. W. "No-Slip Library Machine", Science News Letter 56, 295 (1949).

407. McCormick, E. M. "Bibliography on Mechanized Library Processes", National Science Foundation, Washington, D. C. Apr 1963, 27 p.

408. McCormick, E. M. "Some Observations on Mechanization of Library Processes" in H. P. Luhn [ed]. "Automation and Scientific Communication, Short Papers, Pt. 2", 1963, p. 2.

409. McCormick, E. M. "A Trend in the Use of Computers for Information Processing", Amer. Documentation 13, 182-184 (1962).

410. McCormick, E. M. "Why Computers?", in "Machine Indexing", American U., 1962, p. 220-232.

411. McCulley, W. R. "UNIVAC Compiles a Complete Bible Concordance", Systems 20, 22-23 (1956).

412. McGee, L. L., W. J. Holliman, A. Z. Loren, Jr. and G. D. Adams, "Compilation and Computer Updating of a Medical Sciences Thesaurus", in H. P. Luhn [ed]. "Automation and Scientific Communication, Short Papers, Pt. 2", 1963, p. 347-348.

413. Meetham, A.R. "Preliminary Studies for Machine Generated Index Vocabularies", Language and Speech 6, 22-36 (1963).

414. Melton, J., G. Putnam, W. Goffman and C. Hespen, "Automatic Processing of Metallurgical Abstracts for the Purpose of Information Retrieval", Prog. rept. NSF G-24488, Center for Documentation and Communication Research, Western Reserve University, Cleveland, Ohio, 10 Jan 1963, 15 p.

415. Mersel, J. and S.B. Smith, "Center for Text in Machine-Usable Form", a feasibility study under contract NSF-C320 with the National Science Foundation, TRW Computer Division, Thompson Ramo Wooldridge, Inc., Canoga Park, Cal. 28 Feb 1964, 103 p.

416. Metcalfe, J. "Information Indexing and Subject Cataloging: -- Alphabetical, Classified, Coordinate, Mechanical", Scarecrow Press, New York, 1957, 338 p.

417. Meyer-Uhlenried, K.H. and G. Lustig, "Analysis, Indexing and Correlation of Information", in H. P. Luhn [ed]. "Automation and Scientific Communication, Short Papers, Pt. 2", 1963, p. 229.

418. Mikhailov, A.I. "Problems of Mechanization and Automation of Information Work", Rev. Int. Doc. 29, 49-56 (1962).

419. Miller, E., D. Ballard, J. Kingston and M. Taube, "Conventional and Inverted Grouping of Codes for Chemical Data", in "Proceedings of the International Conference on Scientific Information", 1959, Vol I, p. 671-685.

420. Mimosa Frenk Foundation for Applied Neurochemistry, "KWIC Index to Neurochemistry", Amsterdam, the Netherlands, Aug 1961, 123 p.

421. Montgomery, C. and D.R. Swanson, "Machine-Like Indexing By People", Amer. Documentation 13, 359-366 (1962).

422. Mooers, C.N. "The Next Twenty Years in Information Retrieval: Some Goals and Predictions", Amer. Documentation 11, 229-236 (1960).

423. Mooers, C.N. "Summary of Lectures No. 1 and No. 2", presented at NATO Advanced Study Institute on Automatic Document Analysis, Venice, 7-20 July 1963, 4 p.

424. Mooers, C.N. "Summary of Lectures Nos. 3, 4 and 5", presented at NATO Advanced Study Institute on Automatic Document Analysis, Venice, 7-20 July 1963, 7 p.

425. Moss, R. "How Do We Classify?", ASLIB Proc. 14, 33-42 (1962).

426. National Physical Laboratory, "International Conference on Machine Translation of Languages and Applied Language Analysis (1961)", proceedings of the conference held at the National Physical Laboratory, 5-8 Sep 1961: National Physical Laboratory Symposium No. 13, Vol I, Her Majesty's Stationery Office, London, 1962, p. 1-401.

427. National Physical Laboratory, "International Conference on Machine Translation of Languages and Applied Language Analysis (1961)", proceedings of the conference held at the National Physical Laboratory, 5-8 Sep 1961: National Physical Laboratory Symposium No. 13, Vol II, Her Majesty's Stationery Office, London, 1962, p. 403-747.

428. National Physical Laboratory, "Mechanisation of Thought Processes", proceedings of a symposium held at the National Physical Laboratory, 24-27 Nov 1958: National Physical Laboratory Symposium No. 10, Vol I, Her Majesty's Stationery Office, London, 1959, p. 1-531.

429. National Physical Laboratory, "Mechanisation of Thought Processes", proceedings of a symposium held at the National Physical Laboratory, 24-27 Nov 1958: National Physical Laboratory Symposium No. 10, Vol II, Her Majesty's Stationery Office, London, 1959, p. 533-980.

430. National Science Foundation, "Current Research and Development in Scientific Documentation", No. 1, July 1957, 54 p; No. 3 (NSF-58-33), Oct 1958, 76 p; No. 4 (NSF-59-28), Apr 1959, 85 p; No. 5 (NSF-59-54), Oct 1959, 102 p; No. 6 (NSF-60-25) May 1960, 130 p; No. 10 (NSF-62-20) May 1962, 382 p; No. 11 (NSF-63-5) Nov 1962, 440 p. U.S. Government Printing Office, Washington, D.C.

431. Needham, R.M. "A Method for Using Computers in Information Classification", in C.M. Popplewell, "Information Processing 1962", 1963, p. 284-287.

432. Needham, R.M. "The Place of Automatic Classification in Information Retrieval", presented at NATO Advanced Study Institute on Automatic Document Analysis, Venice, 7-20 July 1963, Rept. no. ML 166, Cambridge Language Research Unit, Cambridge, England, 1963, 8 p.

433. Needham, R.M. "Practical Techniques and Experiments", (Preprint abstract). NATO Advanced Study Institute on Automatic Document Analysis, Venice, July 1963.

434. Needham, R.M. "Research on Information Retrieval Classification and Grouping", Rept. no. ML 149, Cambridge Language Research Unit, Cambridge, England, 1961, lv.

435. Needham, R.M. "The Theory of Clumps, II", Rept. no. ML 139, Cambridge Language Research Unit, Cambridge, England, Mar 1961, 48 p.

436. Needham, R.M., A.H.J. Miller and K. Spärck-Jones, "The Information Retrieval System of the Cambridge Language Research Unit", Rept. no. ML 109, Cambridge, Language Research Unit, Cambridge, England, 1960, 63 p.

437. Netherwood, D.B. "Logical Machine Design: A Selected Bibliography". IRE Transactions on Electronic Computers, EC-7, 155-178 (1958). Corrections to above in IRE Transactions on Electronic Computers, EC-7, 250 (1958).

438. Newbaker, H.R. and T.R. Savage, "Selected Words in Full Title: A New Program for Computer Indexing", in H.P. Luhn [ed]. "Automation and Scientific Communication, Short Papers, Pt. 2", 1963, p. 87-88.

439. Newman, S.M., R.W. Swanson and K.C. Knowlton, "A Notation System for Transliterating Technical and Scientific Texts for Use in Data Processing Systems", in A. Kent, "Information Retrieval and Machine Translation", 1960, p. 345-376.

440. Northrop Corp. "Outline of a Plan for the Development of an Intelligence Language to Facilitate the Automatic Processing of Complex Conceptual Information", Rept. no. NB 60-152, Hawthorne, Cal. 6 June 1960, 12 p.

441. Nugent, W.R. "A Machine Language for Document Transliteration", preprint, 14th ACM annual meeting, Cambridge, Mass. Sep 1959.

442. O'Connor, J. "Ledley's Tabledex Index: Description and Possible Improvements", Appendix A, "The Scan-Column Index", 1960, p. 55-88.

443. O'Connor, J. "Mechanized Indexing Methods and Their Testing", Institute for Scientific Information, Philadelphia, Pa. 1963, 29 p.

444. O'Connor, J. "Mechanized Indexing: Some General Remarks and Some Small-Scale Empirical Results", (based on a talk given 16 Nov 1960 at the Office of Naval Research Data Processing Seminar, Washington) Institute for Cooperative Research, University of Pennsylvania, Philadelphia, Pa. 31 p.

445. O'Connor, J. "Mechanized Indexing Studies of MSD Toxicity, Part I", Institute for Scientific Information, Philadelphia, Pa. undated, 15 p.

446. O'Connor, J. "The Possibilities of Document Grouping for Reducing Retrieval Storage Size and Search Time", in A. Kent [ed]. "Information Retrieval and Machine Translation", 1960, p. 237-279.

447. O'Connor, J. "Some Remarks on Mechanized Indexing and Some Small Scale Empirical Results", in "Machine Indexing", American U., 1962, p. 266-279.

448. O'Connor, J. "Some Suggested Mechanized Indexing Investigations Which Require No Machines", Institute for Cooperative Research, Univ. of Pennsylvania, Philadelphia, Pa., 1960, 19 p. Also in Amer. Documentation 12, 198-203 (1961).

449. O'Connor, J. "The Scan-Column Index: A Book Form Coordinate Information Retrieval System", Remington Rand-UNIVAC, Philadelphia, Pa. Feb 1960, 52 p. Abridged in Amer. Documentation 13, 204-209 (1962).

450. Office of Technical Services, U.S. Department of Commerce, "Keywords Index to U. S. Government Technical Reports" (Permuted Title Index) Vol 2, no. 3, 15 July 1963.

451. Ohlman, H. "Chronological Bibliography of Permutation Indexing", System Development Corp., Santa Monica, Cal. 1960, lv.

452. Ohlman, H. "Mechanical Indexing: Historical Development, Techniques, and Critique", paper presented at the Annual Meeting of the American Documentation Institute, Berkeley, Cal. Oct 1960, 32 p.

453. Ohlman, H. "Permutation Indexing: Multiple-Entry Listing on Electronic Accounting Machines", System Development Corp., Santa Monica, Cal. unpublished, 5 Nov 1957.

454. Olmer, J. and R. Rich, "A Flexible Direct File Approach to Information Retrieval-Text Edit, Search or Select and Print on an IBM 1401", in "Proceedings Fall Joint Computer Conference", 1963, p. 173-182.

455. Olney, J.C. "Building a Concept Network to Retrieve Information from Large Libraries: Part I". Rept. no. TM-634, System Development Corp., Santa Monica, Cal. 26 Jan 1962, 13 p.

456. Olney, J.C. "Constructing an Artificial Language for Mechanical Indexing", Field Note FN-5119, System Development Corp., Santa Monica, Cal. 2 Sep 1961, 10 p.

457. Olney, J.C. "Feat, An Inventory Program for Information Retrieval", FN-4018, System Development Corp., Santa Monica, Cal. 25 July 1960, 7 p.

458. Olney, J.C. "Library Cataloging and Classification", Rept. no. TM-1192, System Development Corp., Santa Monica, Cal. 29 Apr 1963, 53 p.

459. Oswald, V.A., Jr., et al. "Automatic Indexing and Abstracting of the Contents of Documents", prepared for Rome Air Development Center, Air Research and Development Command, USAF, RADC-TR-59-208, Planning Research Corp., Los Angeles, Cal. 31 Oct 1959, p. 5-34, 59-133.

460. Painter, A.F. "An Analysis of Duplication and Consistency of Subject Indexing Involved in Report Handling at the Office of Technical Services, U.S. Department of Commerce", Office of Technical Services, Washington, D.C. 1963, 135 p.

461. Painter, J.A. "Computer Preparation of a Poetry Concordance", Comm. Assoc. Computing Machinery 3, 91-95 (1960).

462. Papier, L.S. "Reliability of Scientists in Supplying Titles: Implications for Permuted Title Indexing", ASLIB Proc. 15, 333-337 (1963).

463. Parker, R.H. "Mechanical Aids in College and University Libraries", A. L. A. Bull. 32, 818-819 (1938).

464. Parker-Rhodes, A.F. "Contributions to the Theory of Clumps. The Usefulness and Feasibility of the Theory", Rept. no. ML 138, Cambridge Language Research Unit, Cambridge, England, Mar 1961, 34 p.

465. Parker-Rhodes, A. F. and R. M. Needham, "The Theory of Clumps", Rept. no. ML 126, Cambridge Language Research Unit, Cambridge, England, Feb 1960, 1v.

466. Parkins, P. V. "Approaches to Vocabulary Management in Permuted Title Indexing of Biological Abstracts", in H. P. Luhn [ed]. "Automation and Scientific Communication, Short Papers, Pt. 1", 1963, p. 27-28.

467. Parrish, S. M. [ed]. "A Concordance to the Poems of Matthew Arnold", Cornell University Press, Ithaca, New York, 1959, 1v.

468. Parrish, S. M. "Problems in the Making of Computer Concordances", Studies in Bibliography, U. of Virginia 15, 1-14 (1962).

469. Parsons, E. A. "The Alexandrian Library: Glory of the Hellenic World; Its Rise, Antiquities, and Destructions", Elsevier Press, New York, 1952, 468 p.

470. Peakes, G. L., A. Kent and J. W. Perry [eds]. "Progress Report in Chemical Literature Retrieval", Interscience Publishers, New York, 1957, 217 p.

471. Perry, J. W. "Subject Matter Analysis and Coding - Some Fundamental Considerations", in R. W. Casey, et al [eds]. "Punched Cards: Their Applications to Science and Industry", 1958, 697 p.

472. Pevzner, B. R. and N. I. Styazhkin, "A Method of Special Abstracting", Translation of Proceedings of the Conference on Information Processing, Machine Translation, and Automatic Character Recognition, Moscow, Inst. of Scientific Information, Academy of Sciences USSR, No. 6, 1961, p. 1-14, JPRS-13057, 21 Mar 1962, 22 p.

473. "Physindex", Series A. Physique des gaz ionises et fusion thermonucléaire controlée, 1, no. 2. Commissariat a l'Energie Atomique, 42, Centre d'Etudes Nucléaires de Saclay, Gif-sur-Yvette, France (1963).

474. Plath, W. "Automatic Sentence Diagramming", in National Physical Laboratory, "International Conference on Machine Translation", Symposium No. 13, Vol I, 1962, p. 175-193.

475. Pool, I. D. "Trends in Content Analysis", Illinois Press, Urbana, Ill. 1959, 243 p.

476. Popplewell, C. M. [ed]. "Information Processing 1962", Proceedings of IFIP Congress, Munich, 27 Aug - 1 Sep 1962, North-Holland Publishing Co., Amsterdam, 1963, 780 p.

477. Powell, S. and W. O. S. Sutherland, Jr. "Techniques for a Subject-Index of 18th Century Journals", Lib. Chron. Univ. of Texas 5, 6-15 (1956).

478. "Preprints of Papers for the International Conference on Scientific Information", National Academy of Sciences-National Research Council, Washington, D. C. 1958, 1v.

479. President's Science Advisory Committee, "Science, Government, and Information", U. S. Government Printing Office, Washington, D. C. 10 Jan 1963, 52 p.

480. "Proceedings of the International Conference on Scientific Information", National Academy of Sciences-National Research Council, Washington, D. C. 1959, Vol 1, p. 1-813.

481. "Proceedings of the International Conference on Scientific Information", National Academy of Sciences-National Research Council, Washington, D. C. 1959, Vol II, p. 814-1635.

482. "Proceedings of the March AFBMD Conference on Scientific and Technical Information, Washington, March 1960", Air Force Ballistic Missiles Division, Air Research and Development Command, Washington, D. C. 1960, 134 p.

483. "Proceedings, Symposium on Materials Information Retrieval", Tech. Doc. Rept. no. ASD-TDR-63-445, AF Materials Laboratory, Dayton, Ohio, 1962, 159 p.

484. Purto, V.A. "Automatic Abstracting Based on A Statistical Analysis of the Text", Moscow, Institute of Scientific Information, Academy of Sciences USSR, Issue No. 9, p. 1-16. Translation in JPRS-13196, "Foreign Developments in Machine Translation and Information Processing, No. 83, USSR", Joint Publications Research Service, Washington, D.C. 28 Mar 1962.

485. Quemada, B. "L'Inventaire Mecanique des Dictionnaries Bilingues", Bull. d'Information du Laboratoire d'Analyse Lexicologique 4, 13-50 (1961).

486. Quemada, B. "La Mecanisation Dans Les Recherches des Inventaires Lexicologiques", Les Cahiers de Lexicologie 1, 7-46 (1959).

487. Quigley, M. "Library Facts from International Business Machine Cards", Library J. 66, 1065-1067 (1941).

488. Ramo-Wooldridge, Div. of Thompson Ramo Wooldridge, Inc. "The Study for Automatic Abstracting C107-1U12", Los Angeles, Cal. 1961, lv.

489. Ramo-Wooldridge, Div. of Thompson Ramo Wooldridge, Inc. "The Study for Automatic Abstracting C107-1U12", Appendix D, Los Angeles, Cal. 1961, lv.

490. Ramo-Wooldridge, Div. of Thompson Ramo Wooldridge, Inc. "Word Correlation and Automatic Indexing", Progress rept. no. 1, C82-9U9, Los Angeles, Cal. 21 Sep 1959, lv.

491. Ramo-Wooldridge, Div. of Thompson Ramo Wooldridge, Inc. "Word Correlation and Automatic Indexing", Progress rept. no. 2, C82-OU1, Los Angeles, Cal. 21 Dec 1959, lv.

492. Randall, G.E. "Man is Measured by His Horizon", Spec. Libraries 53, 380-381 (1962).

493. Rath, G.J., A. Resnick and T.R. Savage, "The Formation of Abstracts by the Selection of Sentences", Research rept. no. RC-184. IBM Research Center, Yorktown Heights, N.Y. 29 June 1959. Also in Amer. Documentation 12, 139-143 (1961) Part 1. "Sentence Selection by Men and Machines", also in IBM "ACSI-matic Auto-Abstracting Project", Vol 3, 1961, p. 111-117.

494. Ray, L.C. "Automatic Indexing and Abstracting of Natural Languages", in E.A. Tomeski [ed]. "The Clarification, Unification and Integration of Information Storage and Retrieval", 1961, p. 85-94.

495. Ray, L.C. "Description of Computer Program for Text Search", in Thompson Ramo Wooldridge, "Word Correlation and Automatic Indexing Phase I: Final Report", Canoga Park, Cal. 30 Apr 1960, Appendix A.

496. Ray, L.C. "Keypunching Instructions for Total Text Input", in "Machine Indexing", American U., 1962, p. 50-57.

497. Reisner, P. "A Machine Stored Citation Index to Patent Literature Experimentation and Planning", in H.P. Luhn [ed]. "Automation and Scientific Communication, Short Papers, Pt. 1", 1963, p. 71-72.

--- "Report to the Secretary of Commerce by the Advisory Committee on the Application of Machines to Patent Office Problems". See U.S. Department of Commerce.

498. Resnick, A. "The Relative Effectiveness of Titles and Abstracts for Notification in a Selective Dissemination System", Science 134, 1004-1006 (1961).

499. Resnick, A. "The Reliability of People in Selecting Sentences", in IBM, "ACSI-matic Auto-Abstracting Project", Final Report, Vol 3, 1961, p. 118-124. Also in Amer. Documentation 12, 141-143 (1961).

500. Ridenour, L. N. "Bibliography in an Age of Science", in Ridenour, et al., "Bibliography in an Age of Science", 1951, p. 5-35.

501. Ridenour, L. N., R. R. Shaw and A. G. Hill, "Bibliography in an Age of Science", University of Illinois Press, Urbana, Ill. 1951, 90 p.

502. Robinson, J. J. "Automatic Parsing and Fact Retrieval: A Comment and Grammar, Paraphrase, and Meaning", Memo RM-4005-PR, The RAND Corp., Santa Monica, Cal. Feb 1964, 51 p.

503. Rodgers, D. J. "A Study of Inter-Indexer Consistency", General Electric Co., Information Systems Operation, Washington, D. C. 29 Sep 1961, 59 p.

504. Rodgers, D. J. "A Study of Intra-Indexer Consistency", General Electric Co., Information Systems Section, Washington, D. C. Jan 1961, 25 p.

505. Ross, R. M. [ed]. "KWIC Index to the Science Abstracts of China", first edition Dec 1960, prepared for the Symposium on the Sciences of Communist China held by AAAS, 26-27 Dec 1960, 154 p.

506. Ruhl, M. J. "Chemical Documents and Their Titles: Human Concept Indexing vs. KWIC-Machine Indexing", unpublished paper presented at the 144th National Meeting, ACS, Los Angeles, Cal. 2 Apr 1963, 12 p.

507. Ruvinschii, J. "Consignes Provisoires Pour La Mise En Diagrammes Des Textes Scientifiques", Rapport GRISA No. 5, Aug 1960, 27 p. JPRS-10367, p. 38.

508. Ruvinschii, J. "Provisional Instructions for Diagramming Scientific Texts", GRISA (Group for Research on Automatic Scientific Information, EURATOM) rept. no. 6, Sep 1960, 14 p. Translated in Foreign Developments in Machine Translation and Information Processing, France, No. 34, Washington. D. C. U. S. Joint Publications Research Service, JPRS-10367.

509. Sabel, C. S. "The Relation Between Completeness and Effectiveness of a Subject Catalogue", in "Proceedings of the International Conference on Scientific Information", 1959, Vol 1, p. 377-380.

510. Salton, G. "Associative Document Retrieval Techniques Using Bibliographic Information". J. Assoc. Computing Machinery 10, 440-457 (1963).

511. Salton, G. "A Combined Program of Statistical and Linguistic Procedures for Automatic Information Classification and Selection", in H. P. Luhn [ed]. "Automation and Scientific Communication, Short Papers, Pt. 1", 1963, p. 53-54.

512. Salton, G. [ed]. "Information Stroage and Retrieval", Scientific rept. no. ISR-1, Computation Laboratory, Harvard University, Cambridge, Mass. 30 Nov 1961, 152 p.

513. Salton, G. [ed]. "Information Storage and Retrieval", Scientific rept. no. ISR-2, Computation Laboratory, Harvard University, Cambridge, Mass. 1 Sep 1962, lv.

514. Salton, G. [ed]. "Information Storage and Retrieval", Scientific rept. no. ISR-3, Computation Laboratory, Harvard University, Cambridge, Mass. 1 Apr 1963, lv.

515. Salton, G. [ed]. "Information Storage and Retrieval", Scientific rept. no. ISR-4, Computation Laboratory, Harvard University, Cambridge, Mass. 1 Aug 1963, lv.

516. Salton, G. "The Manipulation of Trees in Information Retrieval", in Scientific rept. no. ISR-1, Computation Laboratory, Harvard University, 30 Nov 1961, p. II-1 to II-44, AFCRL-62-77. Also in Comm. Assoc. Computing Machinery 5, 103-114 (1962).

517. Salton, G. "Some Experiments in Automatic Indexing Using Citations and Related Information", presented at NATO Advanced Study Institute on Automatic Document Analysis, Venice, 7-20 July 1963.

518. Salton, G. "Some Experiments in the Generation of Word and Document Associations", in "Proceedings Fall Joint Computer Conference 1962", 1962, p. 234-250.

519. Salton, G. "Some Hierarchical Models for Automatic Document Retrieval", in Scientific rept. no. ISR-3, 1963, p. I-1 to I-34, AFCRL-63-134. Also in Amer. Documentation 14, 213-222 (1963).

520. Salton, G. "The Use of Citations as an Aid to Automatic Content Analysis", Sect. III, in "Information Storage and Retrieval", ISR-2, 1 Sep 1962, p. III-1 to III-51.

521. Savage, T.R. "The Preparation of Auto-Abstracts on the IBM 704 Data Processing System", IBM Research Center, Yorktown Heights, N.Y. 17 Nov 1958, 11 p.

522. Scheele, M. [ed]. "Punched-Card Methods in Research and Documentation (With Special Referency to Biology)", Interscience Publishers, Inc., New York, 1961, 274 p. Vol II of Library Science and Documentation, J.H. Shera [ed].

523. Schneider, K. "Funf Jahre KWIC-Indexing nach H. P. Luhn", Nach.für Dok. 14, 200-205 (1963).

524. Schoenbach, U.H. "Citation Indexes for Science", Science 123, 61-62 (1956).

525. Schullian, D.M. "Ancient Medieval and Renaissance Libraries", article on Libraries, Encyl. Amer. Vol 17, The Americana Corp., New York, 1960 edition, p. 353-358.

526. Schultheiss, L.A., D.S. Culbertson and E.M. Heiliger [eds]. "Advanced Data Processing in the University Library", Scarecrow Press, New York, 1962, 388 p.

527. Schultz, C.K. "Editing Author-Produced Indexing Terms and Phrases via a Magnetic-Tape Thesarus and a Computer Program", in H.P. Luhn [ed]. "Automation and Scientific Communication, Short Papers, Pt. 1", 1963, p. 9.

528. Schultz, C.K. "A Generalized Computer Method for Information Retrieval", in Armed Services Technical Information Agency, "Controlling Literature in Automation", 1960, Washington, D.C. p. 107-130. Also in Amer. Documentation 14, 39-48 (1963).

529. Schultz, C.K. "Some Characteristics of an Efficient Retrieval System", J. Chem. Documentation 2, 103-105 (1962).

530. Schultz, C.K., A. Brooks and P. Schwartz, "Optimization and Standardization of Information Retrieval Language and Systems", Technical Status rept. no. 1, Remington Rand UNIVAC, Blue Bell, Pa. 15 Jan 1961, lv.

531. Schultz, C.K. and P.A. Schwartz, "A Generalized Computer Method for Index Production", Amer. Documentation 13, 420-432 (1962).

532. Schultz, C.K. and C.A. Shepherd, "The 1960 Federation Meeting: Scheduling a Meeting and Preparing an Index by Computer", Federation of American Societies for Experimental Biology, Federation Proceedings 19, 682-699 (1960). Also in Med. Documentation 5, 95-105 (1961).

533. Sebeok, T.A. "Computer Research in Psycholinguistics: A Progress Report", prepared for presentation at the National Symposium on Machine Translation, Los Angeles, Cal. Feb 1960, but not included in Proceedings.

534. Sebeok, T.A. "Notes on the Digital Calculator as a Tool for Analyzing Literary Information", Center for Advanced Study in the Behavioral Sciences, Stanford Univ., undated, 17 p. Also in Poetics, Literary Research Institute, Polish Academy of Sciences, Warsaw, 1961.

535. Sebeok, T.A. and V.J. Zeps, "An Analysis of Structured Content, with Application of Electronic Computer Research in Psycholinguistics", Language and Speech 1, 181-193 (1958).

536. Sebeok, T.A. and V.J. Zeps, "Computer Research in Psycholinguistics: Towards an Analysis of Poetic Language", Behavioral Science 6, 365-369 (1961).

537. Sebeok, T.A. and V.J. Zeps, "A Concordance and Thesaurus of Cheremis Poetic Language", Janua Linguarum, Series Major 8, Mouton and Co., The Hague, 1961, 259 p.

538. Sebestyen, G.S. "Decision-Making Processes in Pattern Recognition", ACM Monograph series, The MacMillan Company, New York, 1962, 162 p.

539. Sebestyen, G.S. "Recognition of Membership in Classes", IRE Trans. Information Theory, IT-7, 44-50 (1961).

540. Secrest, B.W. "The IBM Electronic Statistical Machine Applied to Word Analysis of the Dead Sea Scrolls", IBM World Trade Corp., New York, 17 Nov 1958, 4 p.

541. Seidell, A.H. "Citation System for Patent Office", J. Pat. Office Society 31, 554 (1949).

542. Shaw, R.R. "Machines and the Bibliographical Problems of the Twentieth Century", in L.N. Ridenour et al, "Bibliography in an Age of Science", 1951, p. 37-71.

543. Shaw, R.R. "Parameters for Machine Handling of Alphabetic Information", Amer. Documentation 13, 267-269 (1962).

544. Shepard's Citations, Inc. "How to Use Shepard's Citations", Colorado Springs, Colo. 1873. . . . . . to present.

545. Shepherd, C.A. "The Computer-Stored Thesaurus and Its Use in Concept Processing", in "Proceedings of the Fall Joint Computer Conference 1963", 1963, p. 389-395.

546. Sher, I.H. and E. Garfield, "The Genetics Citation Index Experiment", in H.P. Luhn [ed]. "Automation and Scientific Communication, Short Papers, Pt. 1", 1963, p. 63-64.

547. Shera, J.H. "Mechanical Aids in College and University Libraries", Amer. Lib. Assoc. Bull. 32, 818-819 (1938).

548. Shera, J.H., A. Kent and J.W. Perry [eds]. "Documentation in Action", (Proceedings of Conference on the Practical Utilization of Recorded Knowledge Present and Future), Reinhold Publishing Corp., New York, 1956, 471 p.

549. Sherrod, J. "A Progress Report on an Experiment in Semiautomatic Indexing Conducted by the AEC Division of Technical Information Extension", in H.P. Luhn [ed]. "Automation and Scientific Communication, Short Papers, Pt. 2", 1963, p. 215.

550. Shilling, C.W. "Requirements for a Scientific Mission-Oriented Information Center", Amer. Documentation 14, 49-53 (1963).

551. Shilling, C.W. "Status Report on the Biological Sciences Communication Project (BSCP)", in H.P. Luhn [ed]. "Automation and Scientific Communication, Short Papers, Pt. 2", 1963, p. 205-206.

552. Simmons, R. F. "Synthex: Toward Computer Synthesis of Human Language Behavior", in H. Borko [ed]. "Computer Applications in the Behavioral Sciences", 1962, p. 361-393.

553. Simmons, R. F., S. Klein and K. McConlogue, "Co-Occurrence and Dependency Logic for Answering English Questions", Rept. no. SP-1155, System Development Corp., Santa Monica, Cal. 3 Apr 1963, 30 p.

554. Simmons, R. F., S. Klein and K. McConlogue, "Toward the Synthesis of Human Language Behavior", Rept. no. SP-1155, System Development Corp., Santa Monica, Cal. 27 Sep 1961, 15 p. Also in Behavioral Science 7, 402-407 (1962).

555. Simmons, R. F. and K. McConlogue, "Maximum-Depth Indexing for Computer Retrieval of English Language Data", Amer. Documentation 14, 68-73 (1963). Also SDC Doc. SP-775, System Development Corp., Santa Monica, Cal. 10 Apr 1962.

556. Simons, F. W. "Report From the Canadian Patent Office", in U.S. Patent Office, "Second Annual Meeting of ICIREPAT", 1962, p. 31-35.

557. Skaggs, B. and M. Spangler, "Easing the Route to Retrieval with Permuted Indexes", Business Automation 9, 26-29, 60, (1963).

558. Slamecka, V. "Classificatory, Alphabetical, and Associative Schedules as Aids in Coordinate Indexing", Amer. Documentation 14, 223-228 (1963).

559. Slamecka, V. "Indexing Aids", Final Rept. RADC-TDR-62-579, Documentation, Inc., Bethesda, Md. Jan 1963, 33 p.

560. Slamecka, V. and J. Jacoby, "Effect of Indexing Aids on the Reliability of Indexers", Final technical note, RADC-TDR-63-116, Documentation, Inc., Bethesda, Md. June 1963.

561. Slamecka, V. and P. Zunde, "Automatic Subject Indexing from Textual Condensations", in H. P. Luhn [ed]. "Automation and Scientific Communication, Short Papers, Pt. 2", 1963, p. 139-140.

562. Solomonoff, R. J. "On Machines to Learn to Translate Languages and Retrieve Information", Progress rept. ZTB-134, Zator Co., Cambridge, Mass. Oct 1959, 17 p.

563. Spangler, M. "General Bibliography on Information Storage and Retrieval", Technical Information Series, R62-CD2, General Electric Co., Phoenix, Ariz. 11 Mar 1962, Rev. 1 Oct 1962.

564. Spärck-Jones, K. "Mechanized Semantic Classification", Paper 25 in 1961 International Conference on Machine Translation of Languages and Applied Language Analysis, National Physical Laboratory Symposium no. 13, 1962, Vol II, p. 417-435.

565. Spiegel, J. "Mark I Experimental Corpus and Descriptor Set for the Statistical Association Procedures for Message Content Analysis", Information System Language Studies No. 1, Suppl. 1. Rept. no. SR-79, The Mitre Corp., Bedford, Mass. 1 Jan 1963, lv.

566. Spiegel, J., E. Bennett, E. Haines, R. Vicksell and J. Baker, "Statistical Association Procedures for Message Content Analysis", Information System Language Studies No. 1, Rept. no. SR-79, The Mitre Corp., Bedford, Mass. Oct 1962, 55 p.

567. Stevens, M. E. "Availability of Machine-Usable Natural Language Material", in "Machine Indexing", American U., 1962, p. 58-75.

568. Stevens, M. E. "A Machine Model of Recall", in "Information Processing", 1960, p. 309-315. Also preprint no. UNESCO/NS/ICIP/J. 54, 14 p.

569. Stevens, M. E. "Preliminary Results of a Small-Scale Experiment in Automatic Indexing", presented at NATO Advanced Study Institute on Automatic Document Analysis, Venice, 7-20 July 1963.

570. Stevens, M. E. and G. H. Urban, "Training a Computer to Assign Descriptors to Documents: Experiments in Automatic Indexing", to appear in "Proceedings of the Spring Joint Computer Conference, 1964", Spartan Books, Baltimore, Md.

571. Stiles, H. E. "The Association Factor in Information Retrieval", J. Assoc. Computing Machinery 8, 271-279 (1961).

572. Stiles, H. E. "Machine Retrieval Using the Association Factor", in "Machine Indexing", American U., 1962, p. 192-206.

573. Stiles, H. E. "Progress in the Use of the Association Factor in Information Retrieval", unpublished, Washington, D. C. 15 Nov 1962, 13 p.

574. Stone, E. "The Descriptor Word Index Program", Rept. no. FN-6599, System Development Corp., Santa Monica, Cal. 1 June 1962, 13 p.

575. Stone, P. J., R. F. Bales, J. Namenwirth and D. M. Ogilvie, "The General Inquirer: A Computer System for Content Analysis and Retrieval Based on the Sentence as a Unit of Information", Behavioral Science 7, 484-501 (1962).

576. Stone, P. J. and B. Hunt, "The General Inquirer Extended: Automatic Theme Analysis Using Tree Building Procedures", in C. M. Popplewell [ed]. "Information Processing 1962", 1963, p. 337-338.

577. Storm, E. "Some Experimental Procedures for the Identification of Information Content", Section I, in G. Salton [ed]. "Information Storage and Retrieval", Scientific rept. no. ISR-1, 1961, p. I-1 to I-34.

578. "Summary of Discussions - Area 5", in "Proceedings of the International Conference on Scientific Information", 1959, Vol II, p. 1255-1268.

579. Suse, R. E. "The Search of Full Text in Natural Language by Machine Methods", in "Proceedings of Second Annual Meeting of ICIREPAT", 1962, p. 149-171.

580. Swanson, D. R. "Automatic Indexing and Classification", preprint, NATO Advanced Study Institute on Automatic Document Analysis, Venice, 7-20 July 1963, 4 p.

581. Swanson, D. R. "Automatic Title Analysis", paper presented at NATO Advanced Study Institute on Automatic Document Analysis, Venice, 7-20 July 1963 (substantially excerpted from Montgomery and Swanson, "Machine-Like Indexing by People").

582. Swanson, D. R. "An Experiment in Automatic Text Searching, Word Correlation and Automatic Indexing, Phase 1, Final Report", Thompson Ramo Wooldridge, Inc., Canoga Park, Cal. Report C82-OU4, 30 Apr 1960, Reprinted 3 Nov 1960, 36 p., and Appendix 5 p.

583. Swanson, D. R. "Interrogating a Computer in Natural Language", in C. M. Popplewell [ed]. "Information Processing 1962", 1963, p. 288-293.

584. Swanson, D. R. "Library Goals and the Role of Automation", Spec. Libraries 53, 466-471 (1962).

585. Swanson, D. R. "The Nature of Multiple Meaning", in H. P. Edmundson [ed]. "Proceedings of the National Symposium on Machine Translation", 1961, p. 386-393.

586. Swanson, D. R. "Research Procedures for Automatic Indexing", in "Machine Indexing", American U., 1962, p. 281-304.

587. Swanson, D.R. "Searching Natural Language Text By Computer", Science 132, 1099-1104 (1960).

588. Swihart, S.J. and E. Bodie, "An Input System for Automated Library Indexing and Information Retrieval, Including Preparation of Catalog Cards", Rept. no. SCR-317, Sandia Corp., Albuquerque, N. Mex. Mar 1963, lv.

589. Switzer, P. "Vector Images in Document Retrieval", in G. Salton [ed]. "Information Storage and Retrieval", ISR-4, Aug 1963, p. I-1 to I-38.

590. System Development Corp. "Research Directorate Report", Rept. no. TM-530/005/00, Santa Monica, Cal. July 1962, 171 p.

591. Szemere, F. "A Linguistic Investigation Into the Possibilities of Reading Technical Publications", in "Proceedings of Second Annual Meeting of ICIREPAT", 1962, p. 91-98.

592. Taine, S.I. "The Future of the Published Index", in "Machine Indexing", American U., 1962, p. 144-149.

593. Tanimoto, T.T. "An Elementary Mathematical Theory of Classification and Prediction", International Business Machines Corp., New York, Nov 1958, 10 p.

594. Tanimoto, T.T. "The General Problem of Classification and Indexing", in "Machine Indexing", American U., 1962, p. 233-235.

595. Tasman, P. "Index and Concordance Development for Literary Documentation and Information Retrieval", Abstract in Automatic Documentation in Action/ADIA, Preprints, Internationale Arbeitstagung, 9-12 June 1959, (Frankfurt/Main, Germany, 1959, 45 p.), p. 44.

596. Tasman, P. "Indexing the Dead Sea Scrolls by Electronic Literary Data Processing Methods", International Business Machines, World Trade Corp., New York, Nov 1958, 12 p.

597. Tasman, P. "Literary Data Processing", IBM J. Research and Development 1, 249-256 (1957).

598. Taube, M. "Storage and Retrieval of Information by Means of the Association of Ideas", Amer. Documentation 6, 1-18 (1955).

599. Taube, M. and Associates, "Studies in Coordinate Indexing", Documentation, Inc., Washington, D.C. Vol 1, 1953; Vol 2, 1954; Vol 3, 1956; Vol 4, 1957; Vol 5, 1959.

600. Thompson, M. "Automatic Reference Analysis", Section II, in G. Salton [ed]. "Information Storage and Retrieval", Scientific rept. no. ISR-3, 1 Apr 1963, p. II-1 to II-27.

601. Thompson Ramo Wooldridge, "Automatic Abstracting", C107-301, Canoga Park, Cal. 2 Feb 1963, 53 p.

602. Thompson Ramo Wooldridge, "Automatic Thesaurus Compilation", Computer-aided Research in Machine Translation, Progress rept. no. 4, Canoga Park, Cal. 21 Oct 1963, 29 p.

603. Thompson Ramo Wooldridge, "Experiment in Automatic Abstracting of Russian", Progress rept. no. 3, Canoga Park, Cal. June 1963, lv.

604. Thompson Ramo Wooldridge, "Final Report on the Study of Automatic Abstracting", Rept. no. C107-1012, Canoga Park, Cal. Sep 1961, lv.

605. Thorne, J.P. "Automatic Language Analysis", Final Technical rept. RADC-TDR-63-11, Indiana Univ., Bloomington, Ind. 31 Dec 1962, 172 p.

606. Tomeski, E.A. and R. Westcott [eds]. "The Clarification, Unification & Integration of Information Storage and Retrieval", Proceedings of February 23rd, 1961, Symposium Held at the Biltmore, New York City, N.Y. Management Dynamics, New York, 1961, 94 p.

607. Touloukian, Y.S. [ed]. "Retrieval Guide to Thermophysical Properties Research Literature", McGraw-Hill, New York, Vols I, II, III, 1962, 1963.

608. Trachtenberg, A. "Automatic Document Classification Using Information Theoretical Methods", in H.P. Luhn [ed]. "Automation and Scientific Communication, Short Papers, Pt. 2", 1963, p. 349-350.

609. Tritschler, R.J. "A Computer-Integrated System for Centralized Information Dissemination, Storage, and Retrieval", ASLIB Proc. 14, 473-503 (1962).

610. Tritschler, R.J. "Effective Information Searching Strategies Without 'Perfect' Indexing", Rept. no. IBM TR 00. 1032. Presented at 1963 ACM National Conference, Denver, Colo. 27 Aug 1963, 11 p.

611. Tukey, J.W. "The Citation Index and the Information Problem: Opportunities and Research in Progress", Annual Report, Princeton Univ., Princeton, N.J. 1962, 58 p.

612. Tukey, J.W. "Keeping Research in Contact with the Literature: Citation Indexes and Beyond", J. Chem. Doc. 2, 34-37 (1962).

613. Turner, L.D. "The SAPIR Program System of Automatic Processing and Indexing of Reports", Rept. no. UCRL-6523, Lawrence Radiation Laboratory, Univ. of California, Livermore, Cal. 27 May 1961.

614. Turner, L.D. and J.H. Kennedy, "System of Automatic Processing and Indexing of Reports", Rept. no. UCRL-6510, Lawrence Radiation Laboratory, Univ. of California, Livermore, Cal. 12 July 1961, 29 p.

615. Uhr, L. and C. Vossler, "A Pattern Recognition Program That Generates, Evaluates, and Adjusts Its Own Operators", in "Proceedings of the Western Joint Computer Conference", 1961, p. 555-569.

616. Union Carbide, Oak Ridge National Laboratory Libraries, "Key Word Index Laboratory Reports Received Semiannual Index January-June 1963", Oak Ridge, Tenn. 1963.

617. U.S. Atomic Energy Commission, "The Literature of Nuclear Science: Its Management and Use", (Proceedings of a conference held at Division of Technical Information Extension, Oak Ridge, Tenn. 11-13 Sep 1962), U.S. Atomic Energy Commission, Oak Ridge, Tenn. Dec 1962, 398 p.

618. U.S. Atomic Energy Commission, "Research and Development Abstracts of the USAEC", RDA-3, Oak Ridge, Tenn. July-Sep 1962, 46 p.

619. U.S. Congress, Senate Committee on Government Operations, "Documentation, Indexing, and Retrieval of Scientific Information. A Study of Federal and non-Federal Science Information Processing and Retrieval Programs", Senate Doc. No. 113, 86th Congress, 2nd session, 28 June 1960, U.S. Government Printing Office, Washington, D.C. 1961, 283 p.

620. U.S. Department of Commerce, "Report to the Secretary of Commerce by the Advisory Committee on the Application of Machines to Patent Office Problems", Washington, D.C. 22 Dec 1954, 76 p.

621. U.S. Patent Office, "Second Annual Meeting of ICIREPAT". Proceedings of the Technical Sessions at the Patent Office of the Federal Republic of Germany, Munich, 4-6 Sep 1962, Washington, D.C. 1962, 226 p.

622. Vanby, L. "A Minor Devil's Documentation Dictionary", Amer. Documentation 14, 143 (1963).

623. Vandeputte, N. "Traitement sur Ordinateur IBM 1401 Textes Scientifiques Anglais en une Vue d'Études Linguistiques Statistiques", Rept/Doc/Cen/S/AFD-31, Centre d'Etudes Nucléaires de Saclay, Gif-sur-Yvette, France, Oct 1961, 34 p.

624. Veilleux, M. "Permuted Title Word Indexing: Procedures for Man/Machine System", in "Machine Indexing", American U., 1962, p. 77-111.

625. Vertanes, C.A. "Automation Raps at the Door of the Library Catalog", Spec. Libraries 52, 237-242 (1961).

626. Vickery, B.C. "Classification and Indexing in Science", 2nd ed., Academic Press, New York, 1959, 235 p.

627. Wadding, R.V. "Keyword in Context (KWIC) Indexing on the IBM 7090 DPS", Rept. no. 62-825-440, International Business Machines Corp., Owego, N.Y. Sep 1962.

628. Wadding, R.V. "A Key-Word-In-Context Package OL KWIC II", Share distribution no. 1372, International Business Machines Corp., Owego, N.Y. 1962.

629. Walkowicz, J.L. "A Bibliography of Foreign Developments in Machine Translation and Information Processing", NBS Technical Note 193, U.S. Government Printing Office, Washington, D.C. 10 July 1963, 191 p.

630. Warheit, I.A. "Catalogs from Punch Cards", Amer. Documentation 10, 254 (1959).

631. Warheit, I.A. "Evaluation of Library Techniques for the Control of Research Materials", Amer. Documentation 7, 267-275 (1956).

632. Watson, C. "Computer Generation of Word Association Maps for Man-Machine Communication", Rept. no. SP-1153, System Development Corp., Santa Monica, Cal. 25 Mar 1963, 24 p.

--- Weinberg report, see President's Science Advisory Committee, "Science, Government, and Information", 1963.

633. Weinstein, E.A. and J.B. Spry, "Boeing Slip: Computer Maintained Printed Book Catalogs", in H.P. Luhn [ed]. "Automation and Scientific Communication, Short Papers, Pt. 2", 1963, p. 233-234.

634. Welch Medical Library Indexing Project, "Final Report on Machine Methods for Information Searching", Johns Hopkins Univ., Baltimore, Md. 1955, lv.

635. Welt, I.D. "A Combined Indexing-Abstracting System", in "Proceedings of the International Conference on Scientific Information", 1959, Vol I, p. 449-459.

636. Westbrook, J.H. "Identifying Significant Research", Science 132, 1229-1234 (1960).

637. Wheater, R.H. "A Mechanized Information Storage and Retrieval System with the Option for Manual Access", in H.P. Luhn [ed]. "Automation and Scientific Communication, Short Papers, Pt. 2", 1963, p. 185-186.

638. White, H.S. "The IBM DSD Technical Information Center--A Total Operating System Approach Combining Traditional Library Features and Mechanized Computer Processing", in H.P. Luhn [ed]. "Automation and Scientific Communication, Short Papers, Pt. 2", 1963, p. 287-288.

639. White, S.P. and J. Walsh, "A Computer Library's Approach to Information Retrieval", Spec. Libraries 54, 345-349 (1963).

640. Wightman, J. P. "Chemical Titles as an Aid to Current Chemical Literature", J. Chem. Documentation 1, 16-17 (1961).

641. Wilkinson, W. A. "A Machine-Produced Book Catalog: Why, How and What Next?", Spec. Libraries 54, 137-143 (1963).

642. Williams, J. H., Jr. "A Discriminant Method for Automatically Classifying Documents", in "Proceedings of the Fall Joint Computer Conference 1963", 1963, p. 161-166.

643. Williams, T. M. "From Text to Topic in Mechanized Searching Systems", in H. P. Edmundson [ed]. "Proceedings of the National Symposium on Machine Translation", 1961, p. 358-362.

644. Williams, T. M. "Translating from Ordinary Discourse into Formal Logic--A Preliminary Systems Study", Scientific report, Tech. Note no. AFCRC-TN-56-770, ACF Industries, Inc., Alexandria, Va. Sep-Nov 1956, 110 p.

645. Wilson, R. "Computer Retrieval of Case Law", Southwestern Law J. 16, 409-438 (1962).

646. Wisbey, R. A. "Concordance Making By Electronic Computer: Some Experiences with the Wiener Genesis", Modern Language Review 57, 161-172 (1962).

647. Wisbey, R. A. "Mechanization in Lexicography", in "Freeing the Mind", 1962, p. 218.

648. "With the Masters: Herman Hollerith", Systems and Procedures J. 14, 18-24 (1963).

649. Wood, G. C. "Biological Subject-Indexing and Information Retrieval by Means of Punched Cards", Spec. Libraries 47, 26-31 (1956).

650. Wyllys, R. E. "Automatic Analysis of the Contents of Documents Part I: Historical Review", Field Note FN-6089, System Development Corp., Santa Monica, Cal. 7 Dec 1961, 16 p.

651. Wyllys, R. E. "Automatic Analysis of the Contents of Documents Part II: Document Searches and Condensed Representations", Field Note FN-6170, System Development Corp., Santa Monica, Cal. 10 Jan 1962, 26 p.

652. Wyllys, R. E. "Document Searches and Condensed Representations", in "Joint Man-Computer Indexing and Abstracting", Mitre SS-13, 1962, p. 37-60. Also SDC Doc. SP-804, System Development Corp., Santa Monica, Cal. 1 May 1962.

653. Wyllys, R. E. "Research in Techniques for Improving Automatic Abstracting Procedures", Rept. no. TM-1087/000/01, System Development Corp., Santa Monica, Cal. 19 Apr 1963, 30 p.

654. Yakushin, B. J. "Algorithmic Method of Discriminating Subject Concepts for Index Compilation (Method of Nomenclator Pairs)", in Nauchno-Tekhnicheskaya Informatsiya (Scientific Technical Information), No. 7, Moscow, 1963, p. 12-20. Translation in JPRS:21, 695, "Foreign Developments in Machine Translation and Information Processing, No. 141", Joint Publications Research Service, Washington, D. C. 1 Nov 1963, p. 16-49.

655. Yngve, V. H. "COMIT as an IR Language", Comm. Assoc. Computing Machinery 5, 19-28 (1962).

656. Yngve, V. H. "Computer Programs for Translation", Scientific American 20, 68-76 (1962).

657. Yngve, V. H. "The Feasibility of Machine Searching of English Texts", in "Proceedings of the International Conference on Scientific Information", 1959, Vol II, p. 975-995.

658. Youden, W. W. "Characteristics of Programs for KWIC and Other Computer-Produced Indexes", in H. P. Luhn [ed]. "Automation and Scientific Communication, Short Papers, Pt. 2", 1963, p. 331-332.

659. Youden, W. W. "Index to the Communications of the ACM Volumes 1-5 (1958-1962)", Comm. Assoc. Computing Machinery 6, I-1 to I-32 (1963).

660. Youden, W. W. "Index to the Journal of the Association for Computing Machinery", Vols. 1-10 (1954-1963) J. Assoc. Computing Machinery 10, 583-646 (1963).

661. Zusman, T. S., M. S. Thompson, J. B. Wilson, L. S. Rotolo and T. D. Gomery, "Selected Bibliography of the International Geophysical Year: An Example of Tabledex Formats", National Biomedical Research Foundation, NBR and the Library of Congress, Rept. No. 62071/18100, Washington, D. C. July 1962, 109 p.

662. "1946-1949 Expanded Title Index of U. S. Chemically Related Patents", Information for Industry, Washington, D. C. 1962, lv.

APPENDIX B. PROGRESS AND PROSPECTS IN MECHANIZED INDEXING

A working paper prepared for the Symposium on
Mechanized Abstracting and Indexing, Moscow,
28 September - 1 October 1966

Mary Elizabeth Stevens
National Bureau of Standards
Washington, D. C. 20234

The term mechanized indexing can be interpreted in two different ways: as involving the use of machines to produce indexes once the index entries have been pre-determined manually, or as involving the use of machines to select the index entries as well as to prepare the indexes.

The first interpretation, that of machine compilation of indexes is perhaps best represented by the progressively more sophisticated mechanization used for the production of Index Medicus from manual "shingling", through sequential card camera operations, to the computer-based system using a high-speed phototypesetter, the Photon GRACE 1, 2/. As noted elsewhere in this report, machine capabilities have made practical the preparation of citation indexes. In general, however, machine-compiled indexes work with the results of human intellectual efforts as applied in the subject content analysis of documents. We also find machines used to provide aids to the indexer. Two different tools may be employed to improve the quality of indexing. There are prescriptive aids in the sense of limiting and rigorously defining the scope of index terms to be used, and there are suggestive aids in the sense of provoking ideas about additional terms that might be used.

The first type may involve a mechanized authority list or thesaurus used to normalize proposed index term entries, as has been demonstrated by Schultz 3/ and Schultz and Shepherd 4/ from 1960 onward. The potential value of this technique is indicated by further investigations of Schultz et al 5/ in which it was found that index terms proposed by authors agreed more with terms employed by more than one member of a typical user group than did terms available in the document titles. Another example of developments in the use of a mechanized thesaurus is the system at Lockheed Missiles and Space Division, Palo Alto 6/.

This type of tool is used to check proposed indexing terms against the terms of the system vocabulary, to prescribe choices between synonyms and different levels of specificity, and to supply syndectic devices such as "see also" references. Computer manipulations of thesauri can also be used to diversify search questions and to provide useful groupings of terms previously used in the system. The mechanized thesaurus can thus serve as the second type of aid by suggesting to the human indexer additional terms he might use. In effect, such a thesaurus provides a display of prior term-term, document-term and document-document associations observed in a particular collection, such as was demonstrated in the form of special purpose equipment in Taube's "EDIAC" 7/ and the "ACORN" devices at A. D. Little 8/.

The associational thesaurus can also be used to aid in the resolution of ambiguities of natural language and to provide for updating in the light of changing terminologies or changes in the subject scope of a collection. What are the prospects for automatic updating and revision of a mechanized thesaurus? Luhn 9/ has suggested that a record of the number of times words and groups are looked up would be "an indispensable part of the system for making periodic adjustments based on the usage of words or notions as mechanically established."

Another suggestion for the development of mechanized aids in human indexing procedures has been made by Markus 10/. This is to "explore the possibility of applying programmed teaching to indexing, with or without machines."

Machine-compiled indexes rest upon the efficacy of human indexing and there is increasing reason to doubt that this will be "good enough" for the future. It appears that there is a growing consensus with respect to inadequacies of present scope and coverage of indexing services. Cheydleur 11/ emphasizes that: "The cost of manual classification and abstracting of all the articles in the world's hundred-thousand technical periodicals would be fantastic. The practicality of carrying it out in a coordinated and timely way by manual methods is unrealizable. There is also a pressing need to extend the coverage of a myriad of unpublished working papers. Hence, there is an utter necessity for automatic indexing, abstracting, and summarization by electronic data processors."

Secondly, little confidence can be attached to routine, manual operations to produce subject-content selection indicia for subsequent selection and retrieval of stored documentary items for the following reasons:

1. Wide variations of intra- and inter-analyst consistencies occur in the assignment of content-indicia, even with respect to well-established client-interests and index term vocabularies.

2. Potential clients may or may not be inclined to use the system, regardless of whether or not it provides efficient content-indicator-clue and selection criteria mechanisms.

3. Future queries cannot, in general, be effectively predicted in advance, except for the cases of specific author or title retrieval requests.

The problem of intra-indexer and inter-indexer inconsistency is of special interest because the degree of inconsistency will seriously affect search and retrieval effectiveness and because serious questions are raised with respect to the evaluation of any indexing system in terms of prior or independent human indexing.

With respect to the effect of indexer inconsistency upon subsequent search effectiveness, O'Connor 12/ considers the possibilities of overassignment (i.e., the assignment of indexing terms to an item that a subsequent searcher would not consider pertinent to that item) in the case where a search is specified by index terms A, B and C, each term is over-assigned with ratio 1.0, and assignments and overassignments by the recognition rules are statistically independent: "Then only one eighth of the papers selected by the conjunction of A, B and C would correctly have all three terms."

The complementary disadvantage of missing relevant references on search, because of indexer failure to supply all the appropriate indexing terms that a searcher would have considered relevant to a particular document would imply that, for a three-term query, assuming independence of term-assignments and a consistency level of 50 percent, only 12.5 percent of the documents that the searcher would consider relevant would be retrieved if someone else had indexed these items.

We have previously reported 13/ on the results of 700 simulated 3-term searches based upon both manual and machine indexing of approximately 20 items with respect to a fixed vocabulary of less than 100 allowed descriptors. These results show, that if indexer A assigns to a given document the term "A" as indicative of subject content, then his subsequent chances of retrieving that document with a query for term "A" are 58.4 percent if the item had been indexed by someone other than himself, and 55.8 percent if indexed by an automatic indexing procedure developed at NBS, called SADSACT" (Self-Assigned Descriptors from Self And Cited Titles) 14/. For three-term searches, any one searcher would be able to retrieve 26.4 percent of the items he would consider relevant to his query if they had been indexed by any of the other user-indexers, and 24.7 percent if the items had been indexed by the machine technique.

Tinker 15/ provides evidence on the relationships between inter-indexer inconsistency and retrieval efficiency, assuming that a given indexer is a potential querist, with average chances of retrieval ranging from 6.5 to 36 percent. Additional evidence on the generally unsatisfactory state of manual indexing consistency has been reported as follows:

1. Korotkin and Oliver 16/ report that five psychologists and five non-psychologists indexed 30 items with three descriptors per item. The task was repeated two weeks later with the aid of an alphabetized list of "suggested" descriptors derived from the data acquired in the first session. Mean percent consistency results were as follows:

	Session I	Session II
Group A (Psychologists)	39.0%	53.0%
Group B (Non-psychologists)	36.4%	54.0%

2. Evaluations of relevancy of selected items to a given search request have been explored by Badger and Goffman 17/ as follows: "Each of three evaluators was asked to dissect the output into relevant and non-relevant subsets... A chi-square test was applied to the observed evaluation as compared to those expected if the three evaluators were in complete agreement. The chi-square test of 81.57 was very significant, indicating that there was an absence of agreement."

3. Greer 18/ reports on investigations of the interpersonal agreements between subjects asked to list the search words they would use in posing queries in the field of information storage and retrieval systems. He found "a mean percentage consistency agreement of 26.1 among subjects in stating search words."

4. Hammond 19/ provides a sampling of the use by NASA (National Aeronautics and Space Administration) and DDC (Defence Documentation Center) of a common set of indexing terms to index an identical set of 996 technical reports. In considering 3-term searches against the variant indexing shown in Hammond's tables, sample calculations show a 25-30 percent failure to retrieve potentially relevant items.

5. In terms of intra-indexer consistency, Rodgers 20/ reports that: "A consistency of .59 in selecting words to be indexed on two different occasions is not sufficiently high to give us great confidence in expecting a stable store when human indexers are used."

For these reasons, increasing consideration should be given to the second interpretation of the term "mechanized indexing", that is, to machine generation of index entries, or automatic indexing. This typically involves machine processing of some natural language text, with severe problems of input. The first of several solutions involves use of automatic character recognition techniques to convert printed text to machine-usable form. This approach holds considerable future promise, but there are many current limitations and difficulties.

A second possible solution, manual keyboard operations to produce a machine-useful transcription of a text, is plagued by high costs (i.e., at least $0.01 per word for unverified keypunching), and also by limitations of available time or manpower.

A third alternative is suggested by current developments in computerized typesetting or tape-controlled casting or photocomposition machines. However, while such techniques promise major improvements for the automatic indexing of textual information to be published in the future, little can be done for already available literature, even with respect to the bibliographic citation information alone. Today's difficulties are emphasized by estimates of a cost of 30 million dollars to convert the present Library of Congress catalog to machine-readable form 21/.

Assuming, however, that the input processing problems have been solved, we may ask what machines can do with respect to words in texts, or in portions of texts, that are available in machine-*useful* form? The machines can "read" the words for purposes of shifting and sorting and can copy or reproduce the words in some desired order, as in a machine-prepared concordance. Machines can match input words with words already in store and thus exclude input words from further machine consideration (as by stoplists in KWIC (Keyword-in-Context) and other forms of derivative indexing) or stress certain input words with reference to a selective "inclusion" dictionary.

Next, machines can tabulate and count, so that both absolute and relative word frequency data may be applied to either indexing or search-selection algorithms. Measurements of sequential distances between selected words in the input text may also be applied. Machine look-ups against a master vocabulary can provide automatic supplying of syndectic information, synonym reduction, lexical normalization, generic-specific subsumption, data with respect to previously observed word-word or word-subject co-occurrences. In addition, information can be provided as to the possible syntactic roles of input words.

In the light of such machine capabilities, what can be said of the present state of the art in automatic indexing? Automatic indexing in the sense of machine-prepared indexes that are generated by the automatic extraction and manipulation of keywords, especially from titles, is of course widely used in KWIC indexes such as *Chemical Titles* and many others both in the United States and elsewhere.

Fischer 22/ provides a retrospective view of KWIC indexing concepts, including variants like KWOC (Keyword out of Context) and WADEX (Words and Authors Index to *Applied Mechanics Review*). She stresses the potentialities of linking such extraction indexing to selective dissemination systems and concludes: "Plans for using the 'Echo' satellites to link information centers around the world, in a world wide drive toward immediacy in information dispersion, will surely provide a place for KWIC indexes and for the KWIC concept." Warheit 23/ also reports that consideration is being given to combining selective dissemination systems and KWIC. Fundamental questions remain: *How useful* and *how much used* are KWIC and other machine-generated indexes based upon the extraction of words from a limited portion of the author's own text?

These questions relate to an important distinction between two quite different types of indexing. The distinction is that whereas "derived" indexing takes as index entries the author's own words in the title, the abstract or the full text, in "assignment" indexing an index term, descriptor, subject heading, or classification code is assigned to a document as an indicator of content and the term assigned does not need to be identical with any of the author's own words.

We can report continuing progress in use of derivative indexing techniques such as KWIC, and also in experiments with automatic assignment indexing and automatic subject classification. Timeliness of index production is certainly one of the major virtues of KWIC. A similar timeliness is promised for automatic assignment indexing techniques provided that requirements can be kept sufficiently low with respect both to keystroking and computer processing.

Intermediate results may be achieved by pre-editing, normalization, and post-editing techniques. Manual pre-editing to modify and supplement keywords in title, abstract, or portions of text has been used in permuted title and KWIC-type indexing from the punched card system that began operation in 1952 24/ to the "notation-of-content" system developed for NASA 25/. Kreithen 26/ suggests a combination of derivative and assignment indexing, as follows:

> "The combination of these two automatic indexing methods, whereby a number of indexing terms would be assigned to a document on the basis of its category dependency, and the rest extracted from text, might be a desirable solution."

Automatic assignment indexing, with clue-words in the input textual material used to determine the proper assignments of indexing terms to incoming items, is generally equivalent to automatic classification techniques that assign a single classification category to items, again on the basis of clue-words in the input text, because a minimum cut-off level in the automatic assignment procedure, combined with a sufficiently generic vocabulary, can achieve classificatory as well as indexing results. The present state of the art in automatic assignment indexing and classification is marked by intriguing demonstrations of technical feasibility for the relatively small samples so far investigated. Present difficulties associated with automatic assignment indexing or classification techniques, however, relate to problems of input processing requirements, computational limitations, the special purpose nature of results demonstrated to date, and problems of evaluation.

A listing of automatic classification and assignment indexing experiments as of 1964 is provided in Table 2, pp. 101-103, of the text of this report. To this we should add more recent results of our own as well as additional results reported by O'Connor 27/ and Williams 28, 29/, Dale and Dale 30, 31/, and others.

In the SADSACT method, we start with a "teaching sample" of items representative of our collection, to which indexing terms have previously been assigned. We then derive the statistics of co-occurrences of substantive words in the titles and abstracts of these items with descriptors assigned to them, ending with a vocabulary of clue words weighted with respect to prior co-occurrences with various descriptors with which they have been associated.

Then, for new items, we look up each word of input (typically consisting of 100 words or less: title and up to 10 cited titles, or title and brief abstract, or title and first or last paragraphs) and derive "descriptor-selection-scores" based upon the prior ad hoc word-descriptor associations. The highest ranking descriptors, in terms of the accumulated selection scores, are then assigned, at some appropriate cut-off level, to the new item.

To date, machine first-choice assignments (corresponding to performance figures reported for other automatic classification and indexing experiments) have been checked for 213 test items either against prior DDC indexing or against user evaluations, or both, with 72.3 percent mean overall agreement.

Our most recent results involved 150 test items. Machine assignments of descriptors to items were checked by having up to five actual users of our collection rate the relevance to a given one of 14 descriptors of items whose titles were listed under that descriptor by the machine assignment procedure. A total of 451 pairings of user-relevance-ratings with the machine has now been analyzed, with a mean relevance rating of 74.9 percent. With respect to machine first-choices, there were 206 pairings with 85.4 percent of the machine assignments rated as at least somewhat relevant.

Checks have also been made of SADSACT results as compared to which of these same documents would be directly retrievable if a KWIC or some other title-only index were to be used. For the first 50 machine assignments rated as "highly relevant" in user-evaluations, a check was made to determine whether or not the same item would be retrievable by lookup under the name of the descriptor in a KWIC index. There were 9 such cases, or 18 percent. In 48 percent of the cases, a part of the descriptor name occurred in the document title. For 17 cases, or 34 percent, there were no title words identical with any part of the descriptor name.

One evaluator was also asked to review the titles of 150 test items and to indicate which, if any, he would wish to retrieve under each of 14 descriptors. He requested in all 353 items and 209 of these were retrieved on the basis of the SADSACT assignments, for a recall ratio of 59.2 percent. Of these, 167 had been previously evaluated by the same user for an overall relevance ratio of 81.4 percent.

Summary accounts of automatic classification and assignment indexing experiments have been provided by Schultz 32/ in the form of an "imaginary panel discussion" (in which, hypothetically, Borko, Schultz, and Stevens discuss their respective systems), and by Black 33/ who concludes: "Provided that overall effectiveness is nearly equal, the system that depends less on the human element would clearly seem to be more desirable from a standpoint of reliability and efficiency, and perhaps even from a standpoint of economics as well."

Additional work has been reported by Dale and Dale 30, 31/, Damerau 34/, Dolby et al 35/, Kreithen 26/, O'Connor 27/, and Williams 28, 29/, among others. Borko's 36, 37/ more recent papers on this subject consider problems of reliability and evaluation. He reports comparisons of automatic and manual classifications of 997 psychological abstracts into 11 categories, factor-analytically derived from 65 percent of these abstracts used as source items. He concluded that it was possible to determine that the percentage of agreement between automatic classification and perfectly reliable human classification could reach 67 percent.

O'Connor's 1965 report 12/ provides further promising results of his "machine-like indexing by people" studies and also discussions of other techniques and of difficulties and limitations in automatic indexing experiments to date. Using Merck, Sharp and Dohme indexing data, O'Connor tested additional recognition-of-clue-word rules based on syntactic emphasis, a first sentence and first paragraph measure, a syntactic-distance measure, negations forbidden near clue words, and words naming substances or types of operations being required in close proximity to clue words.

He reports considerable success with these new rules as follows: "The computer rules selected 92% of 180 toxicity papers. Allowing for sampling error, these rules would select between 88 and 95 percent of the toxicity papers. Thus the computer rules would be roughly comparable to, or perhaps superior to, MSD indexers in identifying toxicity papers."

With respect to the difficulties to be observed in automatic indexing experimentation, O'Connor questions the adequacy of samplings of subject specifications, documents, and collections, the size of clue word vocabularies, and the human judgments used as standards in many of the studies that have been made.

The question of sampling adequacy in terms of the representativeness of clue word vocabularies as related to index terms or classification categories may be particularly critical for methods using small teaching samples. Spiegel and Bennett 38/ report that: "There seems to be no simple relation between the size of the corpus and the size of the vocabulary but after a certain point vocabulary size increases very slowly."

Findings by Williams 29/ are encouraging. Working with teaching samples of 35, 70, and 140 items respectively, he reports that in the first 10,000 word tokens processed from the text of 2,700 abstracts 1,800 different word types were encountered but that in the 80,000 to 90,000 range only 255 new types appeared. He found further that "an increase in sample size beyond 140 would not appear to offer any significant increase in classification performance."

Williams found an average correct classification of 62 percent for 474 test items automatically assigned to one of four solid state categories 28/. In other tests, 2,754 solid state abstracts were classified into three primary and three secondary categories, using a computer program capable of handling up to 50 clue words, 10 subject categories, and any number of documents. Performance effectiveness ranged from 62 to 88 percent correct by comparison with the original classifications at the more generic level and from 67 to 92 percent correct at the more specific level.

Further progress in the application of statistical association, clumping and syntactic analysis techniques have also been reported. Statistical association techniques are concerned with correlations and coefficients of similarity assumed to exist between items or objects sharing common properties. In documentary item applications, document-document similarities are calculated for sharings of the same index terms or for common patterns of citing the same references, of being cited by the same other documents, and the like. Word-association techniques include the development of absolute or relative frequencies of co-occurrence in a given set of documents, such as those representative of a specific subject matter field. Various normalizing procedures can be used to remove effects of tendencies for certain words to occur frequently in general. Spiegel and associates 38/ at Mitre Corporation have explored means of normalization to eliminate effects of length of text strings, relative positions of words in a string, and vocabulary size.

Ernst 39/ reports that at Arthur D. Little: "We are ... seeking to provide a working retrieval system which will incorporate associative features. The objective will be to make use of automatically computed index term associations as a basis for detecting and presenting an appropriate list of near-synonyms for the concepts desired by a user --- essentially the automatic generation of a limited thesaurus in response to individual user requests." In Switzer's model 40/, co-occurrence statistics of index terms consisting of words from title or text, author's names, and words and author names from cited titles, are used. Significant probabilities for such co-occurrences are then derived.

Methods that group objects or items in terms of co-occurrence data for their properties or characteristics are involved in the "clumping" techniques as proposed at the Cambridge Language Research Unit. Further investigations into the development of the basic CLRU approach have been conducted at the Linguistic Research Center at the University of Texas, by Dale and others 30, 41/. In this work, simulation of associative document retrieval by computer gave results for 260 computer abstracts, using the same 90 clue words as previously used by Borko: "The recall ratios in the test requests were high (i.e., very few relevant documents were not retrieved); relevance ratios were characteristically smaller (of the order of 10 percent). However, since the output lists are ordered, it is interesting to note that the relevance ratios are significantly much higher in the upper portions of the output lists (roughly between 25 percent and 50 percent in the upper fourth of the output lists), and that recall ratios are still of the order of 50-70 percent."

In 1964 a report of the Astropower Laboratory 42/ outlined a "semantic space screening model" based on the assumptions that keywords or phrases have quantifiable 'values', that by itemizing the keywords in a document sufficient information is obtained for its classification, and that by adding the values for the keywords in a document the pertinence of that document to a particular subject field can be determined. A training sample consisted of 120 abstracts drawn from six subfields of electrical engineering. Results showed successful classification of source items, using four different classification formulas, as ranging from 49 to 96.3 percent. Results with test items ranged from 32.9 to 69.0 percent accuracy.

The automatic indexing, selective dissemination and retrieval system design developed by Ossorio 43/ is based on a system vocabulary subsequently used for the automatic assignment of new items to appropriate locations in a pre-established "classification space". An "attribute space" may also be developed to identify the *kind* of information found in a document, e.g., that it deals with concepts such as weight or physical size rather than with mathematical or space and time concepts.

Both types of "space" in this system are constructed through the use of factor analysis applied to previously established relationships between the terms in the system vocabulary (approximately 1,450 terms) and 49 subject fields and to relevance ratings of attributes with respect to items. Then, "documents are indexed by being assigned a set of coordinates in the classification space by means of the classification Formula and the system vocabulary."

With respect to the use of linguistic techniques in automatic indexing and classification, methods of computational linguistics may be used to derive measures of the probable significance of words in document texts. Damerau 34/ reports experimentation with word subset selection for indexing purposes based upon word occurrence frequencies significantly larger than expected frequencies (following Edmundson and Wyllys, in part), with encouraging results. Findings by Black 33/, Simmons et al 44/, Spiegel and Bennett 38/, and Wallace 45/, among others, suggest the need for continuing investigations in the area of proper discrimination between significant clue words and non-informing words for a particular corpus or collection. Extensive computer processing and analyses such as Dennis 46/ has applied to the legal literature are needed for other subject matter fields. The latter investigator warns that neither raw word frequencies nor the numbers of documents in which a word occurs provide good criteria for distinguishing between trivial or non-informing and significant or informing words. She suggests, instead, that "discrimination increases with the skewness of the word distribution in the file".

Baxendale has suggested that certain types of phrase structures and nominal constructions, as determined by relatively unsophisticated machine syntactic analyses, are useful in revealing appropriate subject-content clues. A recent example is provided by Clarke and Wall 47/: "The hypothesis is that the importance of nominal constructions in selection of index unit candidates places emphasis on the bracketing of all noun phrases." Baxendale's continuing work 48/ further suggests that "through the methods of statistical decision theory it is hoped to formulate quantitative measures that will separate informative index terms from noninformative." Continuing use of syntactic analysis principles is provided as an option in the SMART system (Salton 49/) and possibilities for choosing index terms automatically by syntactic criteria have been explored by Dolby et al 35/.

Closely related to automatic classification or indexing experiments involving linguistic factors are document and word grouping investigations for homograph resolution and subject field identification purposes, such as those of Doyle 50/ and Wallace 45/. Doyle used a Fortran computer program developed by Ward and Hook for iterative automatic groupings of 50 physics and 50 non-physics documents. He was able to show clear-cut separation of two meanings of words such as "force" and "satellite".

A case involving overlaps of word memberships in more than one subject class has been investigated by Wallace 45/. Using word frequency data, he found 48 words in common on the first 100 word-frequency rankings for psychological and computer literature abstracts, with function words predominating. However, using a word rank sum criterion, he was able to separate 50 psychological abstracts from 50 computer abstracts with 78 percent success.

We may thus conclude that the progress and prospects of automatic indexing, as of September 1966, are both provocative and challenging. They are "provocative" because so much in terms of both practical and theoretical accomplishment has already been demonstrated, and "challenging" because so much remains to be done. Further, what remains to be done will in all probability require serious, intensive, and imaginative investigations of a wide variety of questions from the relative usage and acceptability of a KWIC index through possible changes in author and editor practices to the fundamental questions of semantics and human judgment.

Nevertheless, when the results of automatic classification or automatic indexing procedures reach levels of 70 percent or better mean agreement either with human indexers or with potential users evaluating the relevance of items retrieved by such indexing, then the machine methods should be preferred to routine, run-of-the-mill, manual indexing wherever the costs are at least commensurate.

The technical feasibility of achieving such performance levels for a relatively small number of classification categories or a relatively small vocabulary of index terms has already been demonstrated experimentally. There remain unresolved questions of the extent to which it will be possible to apply such techniques to the larger vocabulary requirements and the practical operating considerations in actual collections.

Assuming that we can solve these problems, however, many advantages will accrue. First is the speed with which many items can be indexed --- in a few minutes or hours at most for, say, 10,000 items. Secondly, there are advantages of timeliness and the ease with which an entire collection can be re-indexed or re-classified. A third advantage is the consistency of the machine procedures, especially as compared with the inconsistency to be noted in available data on tests of comparative performance among indexers.

The advantage of ability to re-index quickly, easily, and inexpensively (because most input costs will have been incurred previously) is of major importance in terms of overcoming present barriers to the introduction of improvements in operating systems (since, as Kyle 51/ points out, "The most common reason for not trying new and/or improved techniques of classification and indexing is the difficulty of reclassifying and re-indexing large collections") and in terms of dynamic revision and up-dating (as Borko 37/ emphasizes).

Another advantage, particularly of methods using teaching samples is (as suggested by Mooers as early as 1959 52/), the capability for making assignments of indexing terms in, say, an English language system to items whose texts are written in other languages: French, German, or Russian. This type of advantage can point the way to greater international collaboration in indexing and document control procedures.

A further possibility is suggested by the convergence of automatic indexing techniques based upon teaching samples with adaptive selective dissemination systems and client feedback possibilities, especially those involving "more-like-this!" requests. If we assume a large-scale, multiple-access system with adequate personalized files for the typical client, the common data bank of document identificatory and selection criteria, condensed representations, and full text (if available) can be selectively accessed by him on the basis of automatic indexing generated by his own choice of selection criteria and his own choice of exemplar items for each such criterion.

He may provide a standing-order interest profile with respect to patterns of his own selection criteria, with weighting indications as to relative degrees of interest. Dynamic re-adjustments to standing requests and weightings can be made in accordance both with his responses to notifications and with any "more-like-this" requests received from him. System accounting and usage statistics can provide a feedback warning system as to the adequacy of his selection-criteria set and enable him to initiate re-processing of those documents in the collection likely to be of current interest to him.

We must close, however, with a caveat: if machines have not yet mastered us, neither have we yet the requirements of the machine to the degree of advanced planning that will be required, especially for those information processing operations involving the analysis of content and not merely the manipulation of records: for here we are faced with the great challenges of human communication, human decision-making, and human-problem-solving.

REFERENCES

1. National Library of Medicine, The MEDLARS Story at the National Library of Medicine 74 p. (U.S. Dept. of Health, Education, and Welfare, PHS., Washington, D.C., 1963).

2. Henderson, M.B., J.S. Moats, M.E. Stevens and S.M. Newman, Cooperation, Convertibility and Compatibility Among Information Systems: A Literature Review, NBS Misc. Pub. 276, 140 p. (U.S. Govt. Printing Office, Washington, D.C., June 15, 1966).

3. Schultz, C.K., Editing Author-Produced Indexing Terms and Phrases via a Magnetic-Tape Thesaurus and a Computer Program, in Automation and Scientific Communication, Short Papers, Pt. 1, papers contributed to the Theme Sessions of the 26th Annual Meeting, Am. Doc. Inst., Chicago, Ill., Oct. 6-11, 1963, Ed. H.P. Luhn, p. 9 (Am. Doc. Inst., Washington, D.C., 1963).

4. Schultz, C.K. and C.A. Shepherd, The 1960 Federation Meeting: Scheduling a Meeting and Producing an Index by Computer, Federation of American Societies for Experimental Biology, Federation Proc. 19, 682-699 (1960). Also in Med. Doc. 5, 95-105 (1961).

5. Schultz, C.K., W.L. Schultz and R.H. Orr, Comparative Indexing: Terms Supplied by Biomedical Authors and by Document Titles, Am. Doc. 16, No. 4, 299-312 (Oct. 1965).

6. Drew, D.L., R.K. Summit, R.I. Tanaka and R.B. Whitely, An On-Line Technical Library Reference Retrieval System, Am. Doc. 17, No. 1, 3-7 (Jan. 1966).

7. Taube, M. and Associates, Studies in Coordinate Indexing, Vol. III (Documentation, Inc., Washington, D.C., 1956).

8. Giuliano, V.E., Analog Networks for Word Association, IRE Trans. Military Electron. MIL-7, 221-234 (1963).

9. Luhn, H.P., A Business Intelligence System, IBM J. Res. & Dev. 2, 314-319 (1958).

10. Markus, J., State of the Art of Published Indexes, Am. Doc. 13, No. 1, 15-30 (Jan. 1962).

11. Cheydleur, B.F., Information Retrieval --- 1966, Datamation 7, No. 10, 21-25 (Oct. 1961).

12. O'Connor, J., Automatic Subject Recognition in Scientific Papers: An Empirical Study, J. ACM 12, No. 4, 490-515 (Oct. 1965).

13. Stevens, M.E. and G.H. Urban, Automatic Indexing Using Cited Titles, in Statistical Association Methods for Mechanized Documentation, Symp. Proc., Washington, D.C., 1964, NBS Misc. Pub. 269, Ed. M.E. Stevens et al, pp. 213-215 (U.S. Govt. Printing Office, Washington, D.C., Dec. 15, 1965). Also, unpublished notes for NSF Symp. on Evaluation of Information Selection and Retrieval Systems, Stevens, M.E., Notes on Evaluation Data, Man-Machine Indexing, 1965.

14. Stevens, M. E. and G. H. Urban, Training a Computer to Assign Descriptors to Documents: Experiments in Automatic Indexing, AFIPS Proc. Spring Joint Computer Conf. Vol. 25, Washington, D. C., April 1964, pp. 563-575 (Spartan Books, Baltimore, Md., 1964).

15. Tinker, J. F., Imprecision in Meaning Measured by Inconsistency of Indexing, Am. Doc. 17, No. 2, 96-102 (April 1966).

16. Korotkin, A. L. and L. H. Oliver, The Effect of Subject Matter Familiarity and the Use of an Indexing Aid Upon Inter-Indexing Consistency, 17 p. (General Electric Co., Information Systems Operation, Bethesda, Md., 1964).

17. Badger, G., Jr., and W. Goffman, An Experiment with File Ordering for Information Retrieval, in Parameters of Information Science, Proc. Am. Doc. Inst. Annual Meeting, Vol. 1, Philadelphia, Pa., Oct. 5-8, 1964, pp. 379-381 (Spartan Books, Washington, D.C., 1964).

18. Greer, F. L., Word Usage and Implications for Storage and Retrieval, 74 p. (General Electric Co., Information Systems Operation, Bethesda, Md., 1962).

19. Hammond, W. F., Progress in Automation Among the Large Federal Information Centers, in Second Cong. on the Information System Sciences, held at The Homestead, Hot Springs, Va., Nov. 1964, Ed. J. Spiegel and D. E. Walker, pp. 291-305 (Spartan Books, Washington, D. C., 1965).

20. Rodgers, D. J., A Study of Intra-Indexer Consistency, 25 p. (General Electric Co., Information Systems Operation, Bethesda, Md., 1961).

21. King, G. W., Ed., Automation and the Library of Congress, a survey sponsored by the Council on Library Resources, Inc., 88 p. (Library of Congress, Washington, D. C., 1963).

22. Fischer, M., The KWIC Index Concept: A Retrospective View, Am. Doc. 17, No. 2, 57-70 (April 1966).

23. Warheit, I. A., Dissemination of Information, Lib. Res. & Tech. Serv. 9, 73-89 (Winter 1965).

24. Veilleux, M., Permuted Title Word Indexing: Procedures for a Man-Machine System, in Machine Indexing: Progress and Problems, Proc. Third Inst. on Inf. Storage and Retrieval, Washington, D. C., Feb. 13-17, 1961, pp. 77-111 (The American Univ., Washington, D. C., 1962).

25. Newbaker, H. R. and T. R. Savage, Selected Words in Full Title: A New Program for Computer Indexing, in Automation and Scientific Communication, Short Papers, Pt. 1, papers contributed to the Theme Sessions of the 26th Annual Meeting, Am. Doc. Inst., Chicago, Ill., Oct. 6-11, 1963, Ed. H. P. Luhn, pp. 87-88 (Am. Doc. Inst., Washington, D. C., 1963).

26. Kreithen, A., Vocabulary Control in Automatic Indexing, Data Proc. Mag. 7, No. 2, 60-61 (Feb. 1965).

27. O'Connor, J., Mechanized Indexing Methods and Their Testing, J. ACM 11, No. 4, 437-449 (Oct. 1964).

28. Williams, J. H., Jr., Discriminant Analysis for Content Classification, 272 p. (IBM Corp., Bethesda, Md., Dec. 1965).

29. Williams, J. H., Jr., Results of Classifying Documents with Multiple Discriminant Functions, in Statistical Association Methods For Mechanized Documentation, Symp. Proc., Washington, D. C., 1964, NBS Misc. Pub. 269, Ed. M. E. Stevens et al, pp. 217-224 (U. S. Govt. Printing Office, Washington, D. C., Dec. 15, 1965).

30. Dale, A. G. and N. Dale, Some Clumping Experiments for Associative Document Retrieval, Am. Doc. 16, No. 1, 5-9 (Jan. 1965).

31. Dale, A. G., N. Dale and E. D. Pendergraft, A Programming System for Automatic Classification with Applications in Linguistic and Information Retrieval Research, Rept. No. LRC-64 WTM-Y (Linguistics Research Center, Univ. of Texas, Austin, Oct. 1964).

32. Schultz, C. K., An Imaginary Panel Discussion About Indexing, in Parameters of Information Science, Proc. Am. Doc. Inst. Annual Meeting, Vol. 1, Philadelphia, Pa., Oct. 5-8, 1964, pp. 437-452 (Spartan Books, Washington, D. C., 1964).

33. Black, D. V., Automatic Classification and Indexing, for Libraries?", Lib. Res. & Tech. Serv. 9, No. 1, 35-52 (Winter 1965).

34. Damerau, F. J., An Experiment in Automatic Indexing, Am. Doc. 16, No. 4, 283-289 (Oct. 1965).

35. Dolby, J. L., L. L. Earl and H. L. Resnikoff, The Application of English-Word Morphology to Automatic Indexing and Extracting, Rept. No. M-21-65-1, 1 v. (Lockheed Missiles and Space Co., Palo Alto, Calif., April 1965).

36. Borko, H., Measuring the Reliability of Subject Classification by Men and Machines, Am. Doc. 15, No. 4, 268-273 (Oct. 1964).

37. Borko, H., A Factor Analytically Derived Classification System for Psychological Reports, in Perceptual and Motor Skills, Vol. 20, pp. 393-406, 1965; and Studies on the Reliability and Validity of Factor-Analytically Derived Classification Categories, in Statistical Association Methods For Mechanized Documentation, Symp. Proc., Washington, D. C., 1964, NBS Misc. Pub. 269, Ed. M. E. Stevens et al, pp. 245-258 (U. S. Govt. Printing Office, Washington, D. C., Dec. 15, 1965).

38. Spiegel, J. and E. M. Bennett, A Modified Statistical Association Procedure for Automatic Document Content Analysis and Retrieval, in Statistical Association Methods For Mechanized Documentation, Symp. Proc., Washington, D. C., 1964, NBS Misc. Pub. 269, Ed. M. E. Stevens et al, pp. 245-258 (U. S. Govt. Printing Office, Washington, D. C., Dec. 15, 1965).

39. Ernst, M. L., Evaluation of Performance of Large Information Retrieval Systems, in Second Cong. on the Information System Sciences, held at The Homestead, Hot Springs, Va., Nov. 1964, Ed. J. Spiegel and D. E. Walker, pp. 239-249 (Spartan Books, Washington, D. C., 1965).

40. Switzer, P., Vector Images in Document Retrieval, in Statistical Association Methods For Mechanized Documentation, Symp. Proc., Washington, D. C., 1964, NBS Misc. Pub. 269, Ed. M. E. Stevens et al, pp. 163-171 (U. S. Govt. Printing Office, Washington, D. C., Dec. 15, 1965).

41. Jernigan, R. and A. G. Dale, Set Theoretic Models for Classification and Retrieval, Rept. No. LRC-64-WTM-5, 20 p. (Linguistic Research Center, Univ. of Texas, Austin, Nov. 1964).

42. Astropower Lab., Douglas Aircraft Co., Adaptive Techniques as Applied to Textual Data Retrieval, Rept. No. RADC-TDR-64-206, 219 p. (Douglas Aircraft Co., Newport Beach, Calif., Aug. 1964).

43. Ossorio, P. G., Dissemination Research, Rept. No. RADC-TR-65-314, 75 p. (Rome Air Development Center, Griffiss Air Force Base, New York, Dec. 1965).

44. Simmons, R. F., S. Klein and K. McConlogue, Indexing and Dependency Logic for Answering English Questions, Am. Doc. 15, No. 3, 196-204 (July 1964).

45. Wallace, E. M., Rank Order Patterns of Common Words as Discriminators of Subject Content in Scientific and Technical Prose, in Statistical Association Methods For Mechanized Documentation, Symp. Proc., Washington, D. C., Mar. 17-19, 1964, NBS Misc. Pub. 269, Ed. M. E. Stevens et al, pp. 225-229 (U. S. Govt. Print. Off., Washington, D. C., Dec. 15, 1965).

46. Dennis, S. F., The Construction of a Thesaurus Automatically From a Sample of Text, in Statistical Association Methods For Mechanized Documentation, Symp. Proc., Washington, D. C., Mar. 17-19, 1964, NBS Misc. Pub. 269, Ed. M. E. Stevens et al, pp. 61-148 (U. S. Govt. Print. Off., Washington, D. C., Dec. 15, 1965).

47. Clarke, D. C. and R. E. Wall, An Economical Program for Limited Parsing of English, AFIPS Proc. Fall Joint Computer Conf., Vol. 27, Pt. 1, Las Vegas, Nev., Nov. 30 - Dec. 1, 1965, pp. 307-316 (Spartan Books, Washington, D. C., 1965).

48. Baxendale, P. B., 'Autoindexing' and Indexing by Automatic Processes, Spec. Lib. 56, No. 10, 715-719 (Dec. 1965).

49. Salton, G., An Evaluation Program for Association Indexing, in Statistical Association Methods For Mechanized Documentation, Symp. Proc., Washington, D. C., Mar. 17-19, 1964, NBS Misc. Pub. 269, Ed. M. E. Stevens et al, pp. 201-210 (U. S. Govt. Print. Off., Washington, D. C., Dec. 15, 1965).

50. Doyle, L. B., Some Compromises Between Word Grouping and Document Grouping, in Statistical Association Methods For Mechanized Documentation, Symp. Proc., Washington, D. C., Mar. 17-19, 1964, NBS Misc. Pub. 269, Ed. M. E. Stevens et al, pp. 15-24 (U. S. Govt. Print. Off., Washington, D. C., Dec. 15, 1965).

51. Kyle, B. F., Information Retrieval and Subject Indexing: Cranfield and After, J. Doc. 20, No. 2, 55-69 (June 1964).

52. Mooers, C. N., Information Retrieval Selection Study. Part II: Seven System Models, Rept. No. ZTB-133, Part II, 39 p. (Zator Co., Cambridge, Mass., Aug. 1959).

APPENDIX C: SELECTIVE BIBLIOGRAPHY OF ADDITIONAL REFERENCES*

Aagaard, J.S., BIDAP: A Bibliographic Data Processing Program for Keyword Indexing, Am. Behav. Scient. 10, No. 6, 24-27 (1967).

Abelson, R.P. and C.M. Reich, Implicational Molecules: A Method for Extracting Meaning From Input Sentences, Proc. Int. Joint Conf. on Artificial Intelligence, Washington, D.C., May 7-9, 1969, Ed. D.E. Walker and L.M. Norton, pp. 641-647 (Int. Joint Conf. on Artificial Intelligence, 1969).

Abraham, C.T., S.P. Ghosh and D.K. Ray-Chaudhuri, File Organization Schemes Based on Finite Geometrics, Part II of [D. Lieberman] Studies in Automatic Language Processing, Final Report, pp. 74-105 (IBM Corp., Thomas J. Watson Research Center, Yorktown Heights, N.Y., July 1965).

Abraham, S. and G. Salapina, On Machine Recognition of Synonymy, in Colloquium on the Foundations of Mathematics, Mathematical Machines and Their Applications, Tihany, Hungary, Sept. 1962, pp. 151-156 (Adademiai Kiado, Budapest, 1965).

Ackerman, H.J., J.B. Haglind, H.G. Lindwall and R.E. Maizell, SWIFT: Computerized Storage and Retrieval of Technical Information, J. Chem. Doc. 8, No. 1, 14-19 (Feb. 1968).

Adams, W.M., A Comparison of Some Machine-Produced Indexes, Tech. Rept. No. HIG-65-1, 46 p. (Hawaii Institute of Geophysics, Univ. of Hawaii, Honolulu, Jan. 1965).

Adams, W.M., Relationship of Keywords in Titles to References Cited, Am. Doc. 18, No. 1, 26-32 (Jan. 1967).

Adams, W.M. and L.C. Lockley, Scientists Meet the KWIC Index, Am. Doc. 19, No. 1, 47-59 (Jan. 1968).

Allén, S., Report on Work in Computational Linguistics, paper presented at the Colloque Sur la Mécanisation et l'Automation des Recherches Linguistiques, Prague, Czechoslovakia, June 7-10, 1966.

Altmann, B., A Multiple Testing of the ABC Method and the Development of a Second-Generation Model. Part I: Preliminary Discussions of Methodology. Part II: Test Results and an Analysis of Recall Ratio, Rept. No. TR-1296, 2 Vols. (Harry Diamond Labs., Washington, D.C., 1965).

Altmann, B., A Natural Language Storage and Retrieval (ABC) Method: Its Rationale, Operation, and Further Development Program, J. Chem. Doc. 6, 154-157 (Aug. 1966).

Altmann, B. and W.A. Riessler, Theory, Testing, and Mechanization of the ABC Retrieval System, Am. Doc. 20, No. 1, 6-15 (Jan. 1969).

Amacher, P. and A.C. Norton, A Natural-Language Glossary for an Automated Bibliographic-Retrieval System, Paper III g 1, 33rd Conf. F.I.D. and Int. Cong. on Documentation, Tokyo, Sept. 12-22, 1967, 14 p. (preprint).

---

* This bibliography was compiled by Stella J. Michaels, with the assistance of Betty Anderson and Mary P. King, under the direction of Josephine L. Walkowicz and Mary Elizabeth Stevens.

Apresyan, Yu. D. and K. I. Babitskii, Work on Semantics at the Laboratory of Machine Translation, Moscow State Pedagogical Institute of Foreign Languages, [in Russian] Computational Linguistics, No. 5, 1-18 (1966).

Ariès, P., Fabrication Automatisée d' Index par Phraseclé (Automatic Construction of Indexes by Keywords) Documentaliste, Numero Special, pp. 89-93, 1966.

Ariès, P., Préparation d' Index par un Ordinateur à l'Institut Francais de Recherches Fruitières Outre-mer (I. F. A. C.), (Preparing an Index by Computer at the French Overseas Fruit Research Institute) [in French] Bull. des Bibliotèques de France 12, No. 8, 297-312 (Aug. 1967).

Armitage, J. E. and M. F. Lynch, Articulation in the Generation of Subject Indexes by Computer, J. Chem. Doc. 7, No. 3, 170-178 (Aug. 1967).

Armitage, J. E. and M. F. Lynch, Some Structural Characteristics of Articulated Subject Indexes, Inf. Storage & Retrieval 4, No. 2, 101-111 (June 1968).

Artandi, S., Automatic Book Indexing by Computer, Am. Doc. 15, No. 4, 250-257 (Oct. 1964).

Artandi, S., Automatic Indexing of Drug Information, in Levels of Interaction Between Man and Information, Proc. Am. Doc. Inst. Annual Meeting, Vol. 4, New York, N. Y., Oct. 22-27, 1967, pp. 148-151 (Thompson Book Co., Washington, D. C., 1967).

Artandi, S. A., Keeping up With Mechanization, Lib. J. 90, 4715-4717 (Nov. 1, 1965).

Artandi, S., Mechanical Indexing of Proper Nouns, J. Doc. 19, No. 4, 187-196 (Dec. 1963).

Artandi, S., The Searchers --- Links Between Inquirers and Indexes, Spec. Lib. 57, No. 8, 571-574 (Oct. 1966).

Artandi, S. and S. Baxendale, Model Experiments in Drug Indexing by Computer, Project MEDICO, First Progress Report, 108 p. (Graduate School of Library Service, Rutgers, The State University, New Brunswick, N. J., Jan. 1968).

Artandi, S. and S. Baxendale, Project MEDICO, Third Progress Report, 75 p. (Graduate School of Library Service, Rutgers, The State University, New Brunswick, N. J., 1969).

Artandi, S. and E. H. Wolf, The Effectiveness of Automatically Generated Weights and Links in Mechanical Indexing, Am. Doc. 20, No. 3, 198-202 (July 1969).

Artandi, S. and E. H. Wolf, The Effectiveness of Weights and Links in Automatic Indexing, Project MEDICO, Second Progress Report, 65 p. (Graduate School of Library Service, Rutgers, The State University, New Brunswick, N. J., Nov. 1968).

Atherton, P., Ed., Proc. Second International Study Conf. on Classification Research, Elsinore, Denmark, Sept. 14-18, 1964, 563 p. (Munksgaard, Copenhagen, 1965).

Atherton, P. and H. Borko, A Test of the Factor-Analytically Derived Automated Classification Method Applied to Descriptions of Work and Search Requests of Nuclear Physicists, Rept. No. SP-1905, 15 p. (System Development Corp., Santa Monica, Calif., 1965).

Automatic English Sentence Analysis, Final Scientific Report for 1 Mar. 1964 - 30 June 1965, Rept. No. ILRS-T-11, 650630, 113 p. (IDAMI Language Research Section, Milan, Italy, June 30, 1965).

Baker, F. B., Latent Class Analysis as an Association Model for Information Retrieval, in Statistical Association Methods For Mechanized Documentation, Symp. Proc., Washington, D. C., Mar. 17-19, 1964, NBS Misc. Pub. 269, Ed. M.E. Stevens et al, pp. 149-155 (U.S. Govt. Print. Off., Washington, D. C., Dec. 15, 1965).

Baker, F. T. and J. H. Williams, Research on Automatic Classification, Indexing and Extracting, FRQNCY: A General-Purpose Frequency Program, Annual Progress Report, 60 p. (IBM Corp., Federal Systems Div., Gaithersburg, Md., Aug. 1968).

Ball, G. H., A Comparison of Some Cluster Seeking Techniques, Interim Report, June - Aug. 1966, RADC-TR-66-514, 47 p. (Rome Air Development Center, Griffiss AFB, N. Y., Nov. 1966).

Ball, G. H. and D. J. Hall, A Clustering Technique for Summarizing Multivariate Data, Behav. Sci. 12, 153-155 (Mar. 1967).

Ball, G. H. and D. J. Hall, ISODATA, A Novel Method of Data Analysis and Pattern Classification, 1 v. (Stanford Research Inst., Menlo Park, Calif., April 1965).

Ball, G. H. and D. J. Hall, ISODATA - A Self-Organizing Computer Program for the Design of Pattern Recognition Preprocessing, in Information Processing 1965, Proc. IFIP Congress 65, Vol. 2, New York, N. Y., May 24-29, 1965, Ed. W. A. Kalenich, pp. 329-330 (Spartan Books, Washington, D. C., 1966).

Balz, C. F. and R. H. Stanwood, Comps. and Eds., Literature on Information Retrieval and Machine Translation, 2nd ed., 168 p. (IBM Corp., Gaithersburg, Md., 1966).

Banerji, R., Some Studies in Syntax-Directed-Parsing, in Computation in Linguistics: A Case Book, Ed. P. L. Garvin and B. Spolsky, pp. 76-123 (Indiana Univ. Press, Bloomington, 1966).

Barbash, S. M. and N. V. Dvoretskaya, Nekotorye Rezultaty Eksperimenta po Podgotovke Permutatsionnogo Ukazatelya na Alfavitnykh Schetno-Perforatsionnykh Mashinakh (Some Results of the Compilation of a Permuted Index on Punched-Card Computers), Nauchno-Tekhnicheskaya Informatsiya, Ser. 2, Vo. 6, 21-23 (1968).

Bar-Hillel, Y., Language and Information, 388 p. (Addison-Wessley, Reading, Mass., 1964).

Bar-Hillel, Y., Machine Translation: The End of an Illusion, in Information Processing 1962, Proc. IFIP Congress 62, Munich, Aug. 27 - Sept. 1, 1962, Ed. C. M. Popplewell, pp. 331-332 (North-Holland Pub. Co., Amsterdam, 1963).

Barnes, R. J. Jr., A Nonlinear Variety of Iterative Association Coefficients, (Abstract), in Statistical Association Methods For Mechanized Documentation, Symp. Proc., Washington, D. C., Mar. 17-19, 1964, NBS Misc. Pub. 269, Ed. M. E. Stevens et al, p. 159 (U.S. Govt. Print. Off., Washington, D. C., Dec. 15, 1965).

Barnhard, H.J. and J.M. Long, Computer Auto-Coding, Selecting and Correlating of Radiologic Diagnostic Cases, Am. J. Roentgenology, Radium Therapy & Nuclear Medicine 96, No. 4, 854-863 (April 1966).

Baruch, J.J., Information System Applications, in Annual Review of Information Science and Technology, Vol. 1, Ed. C.A. Cuadra, pp. 255-271 (Interscience Pub., N.Y., 1966).

Bauer, H., Aufstellung eines Thesaurus für Elektrotechnik mit Hilfe von Maschinenlochkarten (Compilation of a Thesaurus for Electrical Engineering with the Aid of Machine Punched Cards) Nachrichten für Dokumentation 18, No. 1, 6-12 (Feb. 1967).

Baxendale, P.B., Autoindexing and Indexing by Automatic Processes, Spec. Lib. 56, No. 10, 715-719 (Dec. 1965).

Baxendale, P.B., Content Analysis, Specification, and Control, in Annual Review of Information Science and Technology, Vol. 1, Ed. C.A. Cuadra, pp. 71-106 (Interscience Pub., N.Y., 1966).

Baxendale, P.B. and D.C. Clarke, Documentation for an Economical Program for the Limited Parsing of English: Lexicon, Grammar, and Flowcharts, Research Rept. RJ 386, 106 p. (IBM Corp., San Jose Research Lab., San Jose, Calif., Aug. 1966).

Becker, J.D., The Modeling of Simple Analogic and Inductive Processes in a Semantic Memory System, Proc. Int. Joint Conf. on Artificial Intelligence, Washington, D.C., May 7-9, 1969, Ed. D.E. Walker and L.M. Norton, pp. 655-668 (Int. Joint Conf. on Artificial Intelligence, 1969).

Bell, C.J., Implicit Information Retrieval, Inf. Storage & Retrieval 4, No. 2, 139-160 (June 1968).

Bernier, C.L., Indexing and Thesauri, Spec. Lib. 59, No. 2, 98-103 (Feb. 1968).

Bernshtein, E.S., The Problem of Automatic Indexing, Nauchno-Tekhnicheskaya Informatsiya, No. 10, 22-25 (1963).

Berul, L., Information Storage and Retrieval: A State-of-the-Art Report, Rept. No. PR 7500-145, 238 p. (Auerbach Corp., Philadelphia, Pa., Sept. 14, 1964).

Bessinger, J.B. Jr., S.M. Parrish and H.F. Arader, Eds., Proc. Literary Data Processing Conf., Yorktown Heights, N.Y., Sept. 9-11, 1964, 329 p. (IBM Corp., White Plains, N.Y., 1964).

Bigelow, J., Rational and Irrational Requirements Upon Information Retrieval Systems, in Second Cong. on the Information System Sciences, Hot Springs, Va., Nov. 1964, Ed. J. Spiegel and D.E. Walker, pp. 85-94 (Spartan Books, Washington, D.C., 1965).

Binford, R.L., A Comparison of Keyword-In-Context (KWIC) Indexing to Manual Indexing, M.S. Thesis, Univ. of Pittsburgh, 1965, 70 p.

Bittner, B.P., I. Browning and L.F. Parman, Computer Interrogation of Efficiently Stored English Language Information Using the N-Tuple Pattern Recognition Technique, in Progress in Information Science and Technology, Proc. Am. Doc. Inst. Annual Meeting, Vol. 3, Santa Monica, Calif., Oct. 3-7, 1966, pp. 399-407 (Adrianne Press, 1966).

Black, D.V., Automatic Classification and Indexing, for Libraries?, Lib. Res. & Tech. Serv. 9, No. 1, 35-52 (Winter 1965).

Black, D.V. and E.A. Farley, Library Automation, in Annual Review of Information Science and Technology, Vol. 1, Ed. C.A. Cuadra, pp. 273-303 (Interscience Pub., N.Y., 1966).

Black, F.S. Jr., A Deductive Question-Answering System, Ph.D. Thesis, Harvard Univ., Cambridge, Mass., June 1964.

Blomgren, G., A. Goodman and L. Kelly, An Experimental Investigation of Automatic Hierarchy Generation, in Information Storage and Retrieval, Rept. No. ISR-11, [G. Salton, Proj. Dir.] pp. VIII-1 to VIII-25 (Cornell Univ., Ithaca, N.Y., June 1966).

Bloomfield, M., Simulated Machine Indexing, Part 1. Physics Abstracts Subject Index Used as a Thesaurus, Spec. Lib. 57, No. 3, 167-171 (Mar. 1966); Part 2. Use of Words from Title and Abstract for Matching Thesauri Headings, Ibid, 57, No. 4, 232-235 (April 1966); Part 3. Chemical Abstracts Index Used as a Thesaurus, Ibid, 57, No. 5, 323-326 (May-June 1966); Part 4. A Technique to Evaluate the Efficiency of Indexing, Ibid, 57, No. 6, 400-403 (July-Aug. 1966).

Bobrow, D.G., Natural Language Input for a Computer Problem Solving System, Project MAC, Rept. No. MAC-TR-1, 128 p. (M.I.T., Cambridge, Mass., Sept. 1964).

Bobrow, D.G., Problems in Natural Language Communication with Computers, Scientific Rept. No. 5, BBN-1439, 19 p. (Bolt, Beranek and Newman, Inc., Cambridge, Mass., Aug. 1966).

Bobrow, D.G., Problems in Natural Language Communication with Computers, IEEE Trans. Human Factors Engr. HFE-8, 52-55 (1967).

Bobrow, D.G., Syntactic Analysis of English by Computer - A Survey, AFIPS Proc. Fall Joint Computer Conf., Vol. 24, Las Vegas, Nev., Nov. 1963, pp. 365-387 (Spartan Books, Baltimore, Md., 1963).

Bobrow, D.G., Syntactic Theories in Computer Implementations, in Automated Language Processing, Ed. H. Borko, pp. 215-251 (Wiley, New York, 1967).

Bobrow, D.G., J.B. Fraser and M.R. Quillian, Automated Language Processing, in Annual Review of Information Science and Technology, Vol. 2, Ed. C.A. Cuadra, pp. 161-186 (Interscience Pub., New York, 1967).

Bobrow, D.G., J.B. Fraser and M.R. Quillian, Survey of Automated Language Processing 1966, Scientific Rept. No. 7, 60 p. (Bolt, Beranek and Newman, Inc., Cambridge, Mass., April 1967).

Bohnert, H. and M. Kochen, The Automated Multilevel Encyclopedia as a New Mode of Scientific Communication, in Some Problems in Information Science, Ed. M. Kochen, pp. 156-160 (The Scarecrow Press, New York, 1965).

Bohnert, H.G. and P.O. Backer, Automatic English-to-Logic Translation in a Simplified Model. A Study in the Logic of Grammar, Final Report No. AFOSR 66-1727, 117 p. (IBM Corp., Yorktown Heights, N.Y., Mar. 1966).

Boldovici, J.A., D. Payne and D.W. McGill, Jr., Evaluation of Machine-Produced Abstracts, Rept. No. RADC-TR-66-150, 1 v. (Rome Air Development Center, Griffiss AFB, New York, May 1966).

Bonner, R. E., On Some Clustering Techniques, IBM J. Res. & Dev. 8, No. 1, 22-32 (Jan. 1964).

Booth, A. D., Characterizing Documents --- A Trial of an Automatic Method, Computers & Automation 14, No. 11, 32-33 (Nov. 1965).

Booth, A. D., Ed., Machine Translation, 529 p. (Wiley, New York, 1967).

Booth, A. D., Mechanical Aids to Linguistics, Chemistry in Canada 17, No. 5, 28-31 (May 1965).

Borillo, A. and J. Virbel, Problèmes Syntaxiques de l'Indexation Automatique de Documents, (Syntactic Problems of Automatic Indexing of Documents) in Deuxième Conférence Internationale sur le Traitement Automatique des Langues, Grenoble, Aug. 1967, Paper No. 26.

Borko, H., Ed., Automated Language Processing, 386 p. (Wiley, New York, 1967).

Borko, H., Design of Information Systems and Services, in Annual Review of Information Science and Technology, Vol. 2, Ed. C. A. Cuadra, pp. 35-61 (Interscience Pub., New York, 1967).

Borko, H., Experimental Studies in Automated Document Classification, Lib. Sci. 3, No. 1, 88-98 (Mar. 1966).

Borko, H., Indexing and Classification, in Automated Language Processing, Ed. H. Borko, pp. 99-125 (Wiley, New York, 1967).

Borko, H., Information Retrieval and Linguistics Project, Tech. Memo. No. TM-676, 10 p. (System Development Corp., Santa Monica, Calif., Jan. 1962).

Borko, H., Measuring the Reliability of Subject Classification by Men and Machines, Am. Doc. 15, No. 4, 268-273 (Oct. 1964).

Borko, H., Research in Computer Based Classification Systems, in Proc. Second International Study Conf. on Classification Research, Elsinore, Denmark, Sept. 14-18, 1964, Ed. P. Atherton, pp. 220-267 (Munksgaard, Copenhagen, 1965).

Borko, H., Studies on the Reliability and Validity of Factor-Analytically Derived Classification Categories, in Statistical Association Methods For Mechanized Documentation, Symp. Proc., Washington, D.C., Mar. 17-19, 1964, NBS Misc. Pub. 269, Ed. M. E. Stevens et al, pp. 245-257 (U.S. Govt. Print. Off., Washington, D.C., Dec. 15, 1965).

Borkowski, C., An Experimental System for Automatic Identification of Personal Names and Personal Titles in Newspaper Texts, Am. Doc. 18, No. 3, 131-138 (July 1967).

Borkowski, C. G., A System for Automatic Recognition of Names of Persons in Newspaper Texts, Research Paper RC-1563, 62 p. (IBM Corp., Thomas J. Watson Research Center, Yorktown Heights, N. Y., Mar. 3, 1966).

Botosh, I., An Experiment in the Automatic Analysis of Esperanto, [in Russian], Computational Linguistics, No. 5, 19-40 (1966).

Bourne, C.P., Evaluation of Indexing Systems, in Annual Review of Information Science and Technology, Vol. 1, Ed. C.A. Cuadra, pp. 171-190 (Interscience Pub., New York, 1966).

Bowles, E.A., Ed., Computers in Humanistic Research, 258 p. (Prentice-Hall, Englewood Cliffs, N.J., 1967).

Bowman, C.M., F.A. Landee, N.W. Lee and M.H. Reslock, A Chemically Oriented Information Storage and Retrieval System. II. Computer Generation of the Wiswesser Notations of Complex Polycyclic Structures, J. Chem. Doc. 8, No. 3, 133-138 (Aug. 1968).

Brandhorst, W.T., Experience with Large-Scale Machine Vocabularies, talk given before the ESRO/ELDO/Eurospace Colloquium on Advanced Documentation Systems and Technology Utilization, Nice, April 24-26, 1967. Session 3: Indexing Languages and Evaluation Techniques, 39 p.

Brannon, P.B., D.F. Burnham, R.M. James and L.A. Bertram, Automated Literature Alerting System, Am. Doc. 20, No. 1, 16-20 (Jan. 1969).

Brauen, T.L., R.C. Holt and T.R. Wilcox, Document Indexing Based on Relevance Feedback, in Information Storage and Retrieval, Rept. No. ISR-14, [G. Salton, Proj. Dir.] pp. X-1 to X-22 (Cornell Univ., Ithaca, N.Y., Oct. 1968).

Brodda, B. and H. Karlgren, Citation Index and Measures of Association in Mechanized Document Retrieval, Paper III e 4, 33rd Conf. F.I.D. and International Congress on Documentation, Tokyo, Sept. 12-22, 1967, 13 p. (preprint).

Brodda, B. and H. Karlgren, Citation Index and Similar Devices in Mechanized Documentation, Rapport Nr. 2 till Kungl. Statskontoret, SKRIPTOR, Stockholm, Sweden, May 27, 1965, 13 p.

Bryant, E.C., Redirection of Research into Associative Retrieval, in Parameters of Information Science, Proc. Am. Doc. Inst. Annual Meeting, Vol. 1, Philadelphia, Pa., Oct. 5-8, 1964, pp. 503-505 (Spartan Books, Washington, D.C., 1964).

Bryant, E.C., A Status Report on Research in Information Retrieval, in Management Information Systems and the Information Specialist, Proc. Symp. Purdue Univ., July 12-13, 1965, Ed. J.M. Houkes, pp. 87-95 (Krannert Graduate School of Industrial Admin., and the University Libraries, Purdue Univ., Lafayette, Ind., 1966).

Bryant, E.C., D.T. Searls and R.H. Shumway, Some Theoretical Aspects of the Improvement of Document Screening by Associative Transformations, Final Rept. AFOSR 66-0171, 48 p. (Westat Research Analysts, Inc., Denver, Colo., Nov. 30, 1965).

Bryant, E.C., D.T. Searls, R.H. Shumway and D.G. Weinman, Associative Adjustments to Reduce Errors in Document Screening, Final Rept. No. 66-301, 78 p. (Westat Research, Inc., Bethesda, Md., Mar. 31, 1967).

Brygoo, P.R. and F. Levery, Essai d'Edition d'un Prototype de "Bulletin Signaletique" Produit par Ordinateur IBM 1401 Equipe d'une Chaine d'Impression a 120 Characters (Prototype Edition of a Trial Issue of the "Bulletin Signaletique" on an IBM-1401 Computer System with a Chain Printer of 120 Characters, 89 p.

Bunker-Ramo Corp., Computer-Aided Research in Machine Translation, Mar. 31, 1967, 277 p.

Burger, J. F., R. E. Long and R. F. Simmons, An Interactive System for Computing Dependencies, Phrase Structures and Kernels, Rept. No. SP-2454/000/01, 28 p. (System Development Corp., Santa Monica, Calif., June 29, 1967).

Busa, R., An Inventory of Fifteen Million Words, Proc. Literary Data Processing Conf., Yorktown Heights, N. Y., Sept. 9-11, 1964, Ed. J. B. Bessinger, Jr., et al, pp. 64-78 (IBM Corp., White Plains, N. Y., 1964).

Caianiello, E. R., On the Analysis of Natural Languages ('Procrustes' Program), Proc. 3rd All Union SSR Congress on Cybernetics, Odessa, Sept. 1965, Rept. No. SPT/Doc/ E. R. C. 1, 12 p.

Caras, G. J., Comparison of Document Abstracts as Sources of Index Terms for Derivative Indexing by Computer, in Levels of Interaction Between Man and Information, Proc. Am. Doc. Inst. Annual Meeting, Vol. 4, New York, N. Y., Oct. 22-27, 1967, pp. 157-161 (Thompson Book Co., Washington, D. C., 1967).

Carlson, A. R., Concept Frequency in Political Text: An Appreciation of a Total Indexing Method of Automated Content Analysis, Behav. Sci. 12, No. 1, 68-72 (Jan. 1967).

Carmody, B. T. and P. E. Jones, Jr., Automatic Derivation of Microsentences, Commun. ACM 9, No. 6, 443-449 (June 1966).

Carney, G. J., Computer-Assisted Index Preparation, in Progress in Information Science and Technology, Proc. Am. Doc. Inst. Annual Meeting, Vol. 3, Santa Monica, Calif., Oct. 3-7, 1966, pp. 329-338 (Adrianne Press, 1966).

Carroll, J. M. and R. Roeloffs, Computer Selection of Keywords Using Word-Frequency Analysis, Am. Doc. 20, No. 3, 227-233 (July 1969).

Cattell, R., A Note on Correlation Clusters and Cluster Search Methods, Psychometrika 9, No. 3, 169-184 (Sept. 1944).

Ceccato, S., Ed., Linguistic Analysis and Programming for Mechanical Translation, Rept. No. RADC-TDR-60-18, 242 p. (Milan Univ., Italy, June 1960).

Center for Text in Machine-Usable Form, Scient. Inf. Notes 5, 1 (1963).

Cheatham, T. E. Jr. and S. Warshall, Translation of Retrieval Requests Couched in a 'Semi-Formal' English-Like Language, Commun. ACM 5, 34-39 (Jan. 1962).

Chen, C.-C. and E. R. Kingham, Subject Reference Lists Produced by Computer, J. Lib. Automation 1, No. 3, 178-197 (Sept. 1968).

Chernyi, A. I., A Criterion for the Semantic Conformity of a Document Retrieval System, translated from Nauchno-Tekhnicheskaya Informatsiya Series 2, No. 9, 17-25 (1967).

Chernyi, A. I., Sintagmaticheskie Otnosheniya Mezhdu Deskriptorami (Symtagmatic Relations Between Descriptors), Nauchno-Tekhnicheskaya Informatsiya Series 2, No. 4, 6-16 (1968).

Cherry, J. W., Computer-Produced Indexes in a Double Dictionary Format, Spec. Lib. 57, No. 2, 107-110 (Feb. 1966).

Cheydleur, B. F., Ed., Colloquium on Technical Preconditions for Retrieval Center Operations, Proc. National Colloquium on Information Retrieval, Philadelphia, Pa., April 24-25, 1964, 156 p. (Spartan Books, Washington, D. C., 1965).

Cheydleur, B. F., Indexing Depth, Retrieval Effectiveness and Time-Sharing, in Electronic Handling of Information: Testing and Evaluation, Ed. A. Kent et al, pp. 151-162 (Thompson Book Co., Washington, D. C., 1967).

Chomsky, N., Aspects of the Theory of Syntax, Special Rept. No. 11, 251 p. (M.I.T. Press, Cambridge, Mass., 1965).

Chomsky, N., Syntactic Structures, 116 p. (Mouton, The Hague, 1957).

Chomsky, N., Three Models for the Description of Language, IRE Trans. Inf. Theory, IT-2, 113-124 (Sept. 1956). Reprinted with corrections in Readings in Mathematical Psychology, Vol. 2, Ed. R. D. Luce et al (Wiley, New York, 1965).

Chonez, N., Permuted Title or Key-Phrase Indexes and the Limiting of Documentalist Work Needs, Inf. Storage & Retrieval 4, No. 2, 161-166 (June 1968).

Chonez, N., Préparation Automatique d'Index sur un Ordinateur IBM 360, paper presented at the Cycle d'Information du Centre d'Information du Matériel et des Articles de Bureau (CIMAB): Incidence de l'Electronique sur la Documentation, Paris, Jan. 17, 1967, Rapport CEA-TP 5056, Gif-sur-Yvette (91), Service Central de Documentation du C. E. A., Jan. 1967, 7 p.

Chu, J. T., Optimal Procedures for Automatic Abstracting, in Colloquium on Technical Preconditions for Retrieval Center Operations, Proc. National Colloquium on Information Retrieval, Philadelphia, Pa., April 24-25, 1964, Ed. B. F. Cheydleur, pp. 103-116 (Spartan Books, Washington, D. C., 1965).

Claridge, P. R. P., Mechanized Indexing of Information on Chemical Compounds in Plants, The Indexer 2, No. 1, 4-19 (Spring 1960).

Clarke, D. C. and R. E. Wall, An Economical Program for Limited Parsing of English, AFIPS Proc. Fall Joint Computer Conf., Vol. 27, Pt. 1, Las Vegas, Nev., Nov. 30 - Dec. 1, 1965, pp. 307-316 (Spartan Books, Washington, D. C., 1965).

Cleverdon, C., The Cranfield Tests on Index Language Devices, Aslib Proc. 19, No. 6, 173-194 (June 1967).

Cleverdon, C. W., Ed., Proc. First Cranfield International Conf. on Mechanized Information Storage and Retrieval Systems, College of Aeronautics, Cranfield, England, Aug. 29-31, 1967, in Inf. Storage & Retrieval 4, No. 2, 83-256 (1968).

Cleverdon, C. W. and M. Keen, Factors Determining the Performance of Indexing Systems, Vol. 2, Test Results, Aslib Cranfield Research Project, Cranfield, England, 1966, 299 p.

Cleverdon, C. W., J. Mills and M. Keen, Factors Determining the Performance of Indexing Systems, Vol. 1, Design, Aslib Cranfield Research Project, Cranfield, England, 1966, 120 p.

Colby, K.M. and D.C. Smith, Dialogues Between Humans and an Artificial Belief System, Proc. Int. Joint Conf. on Artificial Intelligence, Washington, D.C., May 7-9, 1969, Ed. D.E. Walker and L.M. Norton, pp. 319-324 (Int. Joint Conf. on Artificial Intelligence, 1969).

Coles, L.S., An On-Line Question-Answering System with Natural Language and Pictorial Input, Proc. 23rd National Conf., ACM, Las Vegas, Nev., Aug. 27-29, 1968, pp. 157-167 (Brandon/Systems Press, Inc., Princeton, N.J., 1968).

Coles, L.S., Syntax Directed Interpretation of Natural Language, Ph.D. Thesis, Carnegie-Melton Univ., Pittsburgh, Pa., 1967, 167 p.

Collila, R.A. and B.H. Sams, Information for Processing and Retrieving, Commun. ACM 5, 11-16 (Jan. 1962).

Conference Internationale Sur Le Traitement Automatique Des Langues, Proc. COOP, 2nd, Grenoble, Aug. 23-25, 1967, 1 vol. (Grenoble, 1967).

Constantinescu, P., The Classification of a Set of Elements with Respect to a Set of Properties, Computer J. 8, No. 4, 352-357 (Jan. 1966).

Cooper, W.S., Fact Retrieval and Deductive Question-Answering Retrieval Systems, J. ACM 11, No. 2, 117-137 (April 1964).

Cooper, W.S., Is Interindexer Consistency a Hobgoblin?, Am. Doc. 20, No. 3, 268-278 (July 1969).

Cossum, W.E., M.E. Hardenbrook and R.N. Wolfe, Computer Generation of Atom-Bond Connection Tables from Hand-Drawn Chemical Structures, in Parameters of Information Science, Proc. Am. Doc. Inst. Annual Meeting, Vol. 1, Philadelphia, Pa., Oct. 5-8, 1964, pp. 269-275 (Spartan Books, Washington, D.C., 1964).

Cowgill, G.L., Computer Applications in Archaeology, AFIPS Proc. Fall Joint Computer Conf., Vol. 31, Anaheim, Calif., Nov. 14-16, 1967, pp. 331-337 (Thompson Books, Washington, D.C., 1967).

Cox, N.S.M. and M.W. Grose, Eds., Organization and Handling of Bibliographic Records by Computer, 187 p. (Archon Books, Hamden, Conn., 1967).

Coyaud, M., L'Analyse Morphologique en Documentation Automatique, La Traduction Automatique 5, No. 3, 59-62 (Sept. 1964).

Coyaud, M., Manuel de Codage des Mots Francais Pour l'Analyse Automatique, La Traduction Automatique 5, No. 3, 63-68 (Sept. 1964).

Coyaud, M., Le Probleme Des "Auto-Abstracts", invited paper, Symp. on Mechanized Abstracting and Indexing - Papers and Discussion, Moscow, Sept. 28 - Oct. 1, 1966, pp. 25-31, Unesco Doc. No. SC/WS/172, issued Paris, Jan. 12, 1968 (Distribution Limited).

Coyaud, M., Resolution of Lexical Ambiguities in Ophthalmology, in Information Storage and Retrieval, Rept. No. ISR-14, [G. Salton, Proj. Dir.] pp. IV-1 to IV-15 (Cornell Univ., Ithaca, N.Y., Oct. 1968).

Coyaud, M. and N.S. Decauville, L'Analyse Automatique des Documents, (Informatique 1), 148 p. (Mouton, Paris/LaHaye, 1967).

Craft, J.L. and W.B. Strohm, Sentence End Detector for Language Processing, IBM Tech. Disc. Bull., No. 6, 71 p., 1964.

Craig, J.A., S.C. Berezner, H.C. Carney and C.R. Longyear, DEACON: Direct English Access and Control, AFIPS Proc. Fall Joint Computer Conf., Vol. 29, San Francisco, Calif., Nov. 7-10, 1966, pp. 365-380 (Spartan Books, Washington, D.C., 1966).

Cros, R.C., J.C. Gardin and F. Levery, L'Automatisation des Recherches Documentaires - Un Modèle Général ≪LE SYNTOL≫ , 260 p. (Gauthier-Villars, Paris, 1964).

Cuadra, C.A., Ed., Annual Review of Information Science and Technology, Vol. 1, 389 p. (Interscience Pub., New York, 1966).

Cuadra, C.A., Ed., Annual Review of Information Science and Technology, Vol. 2, 484 p. (Interscience Pub., New York, 1967).

Cuadra, C.A., Ed., Annual Review of Information Science and Technology, Vol. 3, 457 p. (Encyclopedia Britannica, Inc., Chicago, Ill., 1968).

Cuadra, C.A., Toward a Scientific Approach to Relevance Judgments, Paper I b 1, 33rd Conf. F.I.D. and Int. Cong. on Documentation, Tokyo, Sept. 12-22, 1967, 12 p. (preprint).

Curtice, R., Discriminating Candidate Single Word Terms, in Papers on Automatic Language Processing, Vol. III, Rept. No. ESD-TR-67-202, pp. 109-123 (Arthur D. Little, Inc., Cambridge, Mass., Feb. 1967).

Curtice, R.M. and P.E. Jones, Distributional Constraints and the Automatic Selection of an Indexing Vocabulary, in Levels of Interaction Between Man and Information, Proc. Am. Doc. Inst. Annual Meeting, Vol. 4, New York, N.Y., Oct. 22-27, 1967, pp. 152-156 (Thompson Book Co., Washington, D.C., 1967).

Dale, A.G., Bases for Improved Information Systems, in Information in the Language Sciences, Proc. Conf. on Information in the Language Sciences, Warrenton, Va., Mar. 4-6, 1966, Sponsored by the Center for Applied Linguistics, Ed. R.R. Freeman et al, pp. 185-195 (American Elsevier Pub. Co., New York, 1968).

Dale, A.G., Indexing and Classification for Interactive Retrieval Systems, Inf. Storage & Retrieval 3, No. 4, 377-383 (Dec. 1967).

Dale, A.G. and N. Dale, Clumping Techniques and Associative Retrieval, in Statistical Association Methods For Mechanized Documentation, Symp. Proc., Washington, D.C., Mar. 17-19, 1964, NBS Misc. Pub. 269, Ed. M.E. Stevens et al, pp. 230-235 (U.S. Govt. Print. Off., Washington, D.C., Dec. 15, 1965).

Dale, A.G. and N. Dale, Some Clumping Experiments for Associative Document Retrieval, Am. Doc. 16, No. 1, 5-9 (Jan. 1965).

Dale, A.G., N. Dale and E.D. Pendergraft, A Programming System for Automatic Classification with Applications in Linguistic and Information Retrieval Research, Rept. No. LRC-64-WTM-4, 19 p. (Linguistics Research Center, Univ. of Texas, Austin, Oct. 1964).

Dale, N., Automatic Classification System User's Manual, Rept. No. LRC-64-TTM-1, 19 p. (Linguistics Research Center, Univ. of Texas, Austin, Nov. 1964).

Dale, N. and W. Tosh, An Experiment in Automatic Linguistic Classification, 17 p. (Univ. of Texas, Austin, July 1965).

Damerau, F. J., An Experiment in Automatic Indexing, IBM Research Rept. (IBM Corp., New York, Feb. 19, 1963). Also in Am. Doc. 16, No. 4, 283-289 (Oct. 1965).

Dammann, J. E., An Experiment in Cluster Detection, Letter to the Editor, IBM J. Res. & Dev. 10, No. 1, 80-88 (Jan. 1966).

Dammers, H. F., Integrated Information Processing and the Case for a National Network, Inf. Storage & Retrieval 4, No. 2, 113-131 (June 1968).

Darlington, J. L., Machine Methods for Proving Logical Arguments Expressed in English, in Information Processing 1965, Proc. IFIP Congress 65, Vol. 2, New York, N. Y., May 24-29, 1965, Ed. W. A. Kalenich, pp. 530-531 (Spartan Books, Washington, D. C., 1966).

Darlington, J. L., Machine Methods for Proving Logical Arguments Expressed in English, Mech. Trans. 8, 41-47 (June-Oct. 1965).

Dattola, R. T., A Fast Algorithm for Automatic Classification, in Information Storage and Retrieval, Rept. No. ISR-14, [G. Salton, Proj. Dir.] pp. V-1 to V-31 (Cornell Univ., Ithaca, N. Y., Oct. 1968).

Dattola, R. T. and D. M. Murray, An Experiment in Automatic Thesaurus Construction, in Information Storage and Retrieval, Rept. No. ISR-13, [G. Salton, Proj. Dir.] pp. VIII-1 to VIII-26 (Cornell Univ., Ithaca, N. Y., Jan. 1968).

Davis, C. H., An Approach to Automated Vocabulary Control in Indexes of Organic Compounds, J. Chem. Doc. 7, No. 3, 131-134 (Aug. 1967).

Dennis, S. F., The Construction of a Thesaurus Automatically From a Sample of Text, in Statistical Association Methods For Mechanized Documentation, Symp. Proc., Washington, D. C., Mar. 17-19, 1964, NBS Misc. Pub. 269, Ed. M. E. Stevens et al, pp. 61-148 (U. S. Govt. Print. Off., Washington, D. C., Dec. 15, 1965).

Dennis, S. F., The Design and Testing of a Fully Automatic Indexing-Searching System for Documents Consisting of Expository Text, in Information Retrieval - A Critical View, Based on Third Annual Colloquium on Information Retrieval, Philadelphia, Pa., May 12-13, 1966, Ed. G. Schecter, pp. 67-94 (Thompson Book Co., Washington, D. C., 1967).

Dennis, S. F., Shall We Put The Law Into The Computer?, Law and Computer Tech. 1, No. 1, 25-28 (Jan. 1968).

Dennis, S. F., Status of American Bar Foundation Research on Automatic Indexing-Searching Computer System, M. U. L. L., 131-132 (Sept. 1965).

Desmond, W. F. and L. A. Barrer, Indexing and Classification: A Selected and Annotated Bibliography, 356 p. (Oak Ridge National Lab., Oak Ridge, Tenn., May 1966).

DeSolla Price, D. J., Statistical Studies of Networks of Scientific Papers, (Abstract), in Statistical Association Methods For Mechanized Documentation, Symp. Proc., Washington, D. C., Mar. 17-19, 1964, NBS Misc. Pub. 269, Ed. M. E. Stevens et al, p. 187 (U. S. Govt. Print. Off., Washington, D. C., Dec. 15, 1965).

DeTollenaere, F., Automation in Lexicology, [in German], Sprachkunde und Informationsverarbeitung, No. 1, 33-44 (1963).

Dieteman, D.C., Using LITE for Research Purposes, AF JAG L. Rev. 8, No. 6, 11-19, 25 (Nov.-Dec. 1966).

Dolby, J.L., The Structure of Indexing --- The Distribution of Structure-Word-Free Back-of-the-Book Entries, in Information Transfer, Proc. Am. Soc. Inf. Sci. Annual Meeting, Vol. 5, Columbus, O., Oct. 20-24, 1968, pp. 65-78 (Greenwood Pub. Corp., New York, 1968).

Dolby, J.L. and H.L. Resnikoff, On the Structure of Written English Words, reprinted from Language 40, No. 2, 167-196 (April-June 1964).

Dolby, J.L., L.L. Earl and H.L. Resnikoff, The Application of English-Word Morphology to Automatic Indexing and Extracting, Rept. No. M-21-65-1, 1 v. (Lockheed Missiles and Space Co., Palo Alto, Calif., April 1965).

Dolezel, L., Ed., Prague Studies in Mathematical Linguistics. Vol. 1: Statistical Linguistics, Algebraic Linguistics, Machine Translation, 240 p. (Univ. of Alabama Press, Tuscaloosa, 1966).

Dollar, C.M., Innovation in Historical Research: A Computer Approach, Computers and the Humanities 3, No. 3, 139-151 (Jan. 1969).

Doudnikoff, B. and A.N. Conner, Jr., Statistical Vocabulary Construction and Vocabulary Control with Optical Coincidence, in Statistical Association Methods For Mechanized Documentation, Symp. Proc., Washington, D.C., Mar. 17-19, 1964, NBS Misc. Pub. 269, Ed. M.E. Stevens et al, pp. 177-180 (U.S. Govt. Print. Off., Washington, D.C., Dec. 15, 1965).

Douglas Aircraft Company, Missile and Space Systems Division, Adaptive Techniques as Applied to Textual Data Retrieval, Final Rept. No. RADC-TDR-64-206, 219 p. (Douglas Aircraft Co., Newport Beach, Calif., Aug. 1964).

Doyle, L.B., Is Automatic Classification a Reasonable Application of Statistical Analysis of Text?, Rept. No. SP-1753, 34 p. (System Development Corp., Santa Monica, Calif., Aug. 31, 1964). Also in J. ACM 12, 473-489 (Oct. 1965).

Doyle, L.B., Breaking the Cost Barrier in Automatic Classification, Rept. No. SP-2516, 62 p. (System Development Corp., Santa Monica, Calif., July 1, 1966).

Doyle, L.B., Re-Expression in Standardized Code to Improve the Automatic Classificability of Text Items, Tech. Memo. TM-2213, 32 p. (System Development Corp., Santa Monica, Calif., Feb. 25, 1965).

Doyle, L.B., Some Compromises Between Word Grouping and Document Grouping, in Statistical Association Methods For Mechanized Documentation, Symp. Proc., Washington, D.C., Mar. 17-19, 1964, NBS Misc. Pub. 269, Ed. M.E. Stevens et al, pp. 15-24 (U.S. Govt. Print. Off., Washington, D.C., Dec. 15, 1965).

Doyle, L.B. and D.A. Blankenship, Technical Advances in Automatic Classification, in Progress in Information Science and Technology, Proc. Am. Doc. Inst. Annual Meeting, Vol. 3, Santa Monica, Calif., Oct. 3-7, 1966, pp. 63-71 (Adrianne Press, 1966).

Duffey, M.M., I.M. Klanberg, S.C. Mahr, L.L. Meier and J.L. Romstad, Computer Indexing of Polymer Patents, J. Chem. Doc. 8, No. 2, 85-88 (May 1968).

Dunphy, D.C., P.J. Stone and M.S. Smith, The General Inquirer: Further Developments in a Computer System for Content Analysis of Verbal Data in the Social Sciences, Behav. Sci. 10, No. 4, 468-480 (Oct. 1965).

Dyson, G.M., Computer Input and the Semantic Organization of Scientific Terms - I, Inf. Storage & Retrieval 3, No. 2, 35-115 (April 1967).

Dzholos, E.M., Automatic Translation of Texts: Algorithms for Determining the Distances Between Words, Nauch.-Tekh. Inf. No. 3, 35-40 (Mar. 1966). Digest in: Automat. Expr. 8, No. 3, 8 (1966).

Dzholos, E.M., O.K. Kuchmii, I.I. Rattseva & Y.A. Shreider, An Algorithm for Automatic Determination of Semantic Coordinates, 17 p. Trans. Of: Nauchno-Tekh. Inf. (USSR) No. 3, 29-34 (1964).

Edmundson, H.P., A Correlation Coefficient for Attributes or Events, in Statistical Association Methods For Mechanized Documentation, Symp. Proc., Washington, D.C., Mar. 17-19, 1964, NBS Misc. Pub. 269, Ed. M.E. Stevens et al, pp. 41-44 (U.S. Govt. Print. Off., Washington, D.C., Dec. 15, 1965).

Edmundson, H.P., New Methods in Automatic Extracting, Tech. Memo. No. TM-3207, 30 p. (System Development Corp., Santa Monica, Calif., Sept. 15, 1966).

Edmundson, H.P. and D.G. Hays, Research Methodology for Machine Translation, in Readings in Automatic Language Processing, Ed. D.G. Hays, pp. 137-147 (American Elsevier Pub., New York, 1966).

Elliott, R.W., A Model for a Fact Retrieval System, Ph.D. Thesis, Computation Center, Univ. of Texas, Austin, May 1965, 159 p.

Enger, I., G.T. Merriman and A.L. Bussemey, Automatic Security Classification Study, Rept. No. RADC-TR-67-472, 66 p. (Rome Air Development Center, Griffiss AFB, N.Y., Oct. 1967).

Engineering Societies Library, Bibliography on Filing, Classification and Indexing Systems, and Thesauri for Engineering Offices and Libraries, ESL Bibliography No. 15, New York, 1966, 37 p.

Fangmeyer, H. and G. Lustig, The EURATOM Automatic Indexing Project, preprint of paper presented at the IFIP Conf., Edinburgh, 1968, 5 p.

Farrell, J., TEXTIR: A Natural Language Information Retrieval System, Tech. Memo. No. TM-2392, 37 p. (System Development Corp., Santa Monica, Calif., May 5, 1965).

Fédération Internationale de la Documentation, 31st Meeting and Congress, Washington, D.C., Oct. 7-16, 1965, 371 p. (Spartan Books, Washington, D.C., 1966).

Feigenbaum, E.A. and J. Feldman, Eds., Computers and Thought, 535 p. (McGraw-Hill, New York, 1963).

Feigenbaum, E.A. and H.A. Simon, Performance of a Reading Task by an Elementary Perceiving and Memorizing Program, Behav. Sci. 8, No. 1, 72-76 (1962).

Feldman, A., D.B. Holland and D.P. Jacobus, The Automatic Encoding of Chemical Structures, J. Chem. Doc. 3, No. 4, 187-188 (Oct. 1963).

Findler, N.V. and W.R. McKinzie, Computers In Behavioral Science, On a Computer Program that Generates and Queries Kinship Structures, Behav. Sci. 14, No. 4, 334-340 (July 1969).

Fischer, M., The KWIC Index Concept: A Retrospective View, Am. Doc. 17, No. 2, 57-70 (April 1966).

Fogl, J., On the Question of Forming a Subject Index of an Information Language for the Purpose of Mechanical Indexing of Information by Means of the System of Headings, Metodika a Technika Informaci Pragne 9, 20-25 (1965).

Ford, J.D. Jr., Automatic Detection of Psychological Dimensions in Psychotherapy Transcripts by Means of Content Words, Rept. No. SP-1220, 14 p. (System Development Corp., Santa Monica, Calif., July 12, 1963).

Forman, B., An Experiment in Semantic Classification, Rept. No. LRC-65-WT-3, 22 p. (Linguistics Research Center, Univ. of Texas, Austin, Dec. 1965).

Foskett, A.C., SLIC Indexing, Libr. World 70, No. 817, 17-19 (July 1968).

Fossum, E.G. and G. Kaskey, Optimization of Information Retrieval Language and Systems, Final Rept. under Contract AF49(638)1194, 87 p. (UNIVAC, Blue Bell, Pa., Jan. 28, 1966).

Foster, D.R., Automatic Sentence Kernelization, in Mathematical Linguistics and Automatic Translation, Rept. No. NSF-15, to the National Science Foundation, Ed. A.G. Oettinger, pp. V-1 to V-29 (Computation Lab., Harvard Univ., Cambridge, Mass., Aug. 1965).

Fox, L., Ed., Advances in Programming and Non-Numerical Computation, 218 p. (Pergamon, New York, 1966).

Freeman, R.R., Evaluation of the Retrieval of Metallergical Document References Using the UDC in a Computer-Based System, Rept. No. AIP/UDC-6, 64 [96] p. (American Institute of Physics, New York, 1968).

Freeman, R.R. and P. Atherton, File Organization and Search Strategy Using the Universal Decimal Classification in Mechanized Reference Retrieval Systems, Rept. No. AIP/UDC-5, 30 p. (American Institute of Physics, New York, Sept. 15, 1967). Also in Mechanized Information Storage, Retrieval and Dissemination, Proc. F.I.D./I.F.I.P. Joint Conf., Rome, Italy, June 14-17, 1967, Ed. K. Samuelson, pp. 122-152 (North-Holland Pub. Co., Amsterdam, 1968).

Freeman, R.R., A. Pietrzyk and A.H. Roberts, Eds., Information in the Language Sciences, Proc. Conf. on Information in the Language Sciences, Warrenton, Va., Mar. 4-6, 1966, Sponsored by the Center for Applied Linguistics, 247 p. (American Elsevier Pub. Co., New York, 1968).

Fried, C. and J.J. Prevel, Effects of Indexing Aids on Indexing Performance, Rept. No. RADC-TR-66-525, 192 p. (Rome Air Development Center, Griffiss AFB, New York, Oct. 1966).

Fried, J.B., B.C. Landry, D.M. Liston, Jr., B.P. Price, R.C. VanBuskirk and D.M. Wachsberger, Index Simulation Feasibility and Automatic Document Classification, Tech. Rept. No. 68-4, 21 p. (Computer and Information Science Research Center, The Ohio State Univ., Columbus, Oct. 1968).

Friedman, J., A Computer System for Transformational Grammar, Commun. ACM 12, No. 6, 341-348 (June 1969).

Frielink, A.B., The Possibilities of Automation in Linguistics, [in German], Sprachkunde und Informationsverarbeitung, No. 1, 10-22 (1963).

Furth, S.E., Automated Information Retrieval --- A State of the Art Report, M.U.L.L., 189-190 (Dec. 1965).

Furth, S.E., Automated Retrieval of Legal Information: State of the Art, Computers & Automation 17, No. 12, 25-28 (Dec. 1968).

Gallizia, A., F. Mollame and E. Maretti, Towards the Automatic Analysis of Natural Language Texts, [translation from Italian original], 17 p. (Advisory Group for Aerospace Research and Development, NATO, Paris, 1966).

Gammon, E., A Probabilistic Method for Phrase Determination, 36 p. (Lockheed Missiles and Space Co., Sunnyvale, Calif., Mar. 1962).

Gardin, J.C., Document Analysis and Information Retrieval, Unesco Bull. Lib. 14, 2-5 (1960).

Gardin, J.C., On Some Reciprocal Requirements of Linguistics and Information Techniques, in Information in the Language Sciences, Proc. Conf. on Information in the Language Sciences, Warrenton, Va., Mar. 4-6, 1966, Sponsored by the Center for Applied Linguistics, Ed. R.R. Freeman et al, pp. 95-103 (American Elsevier Pub. Co., New York, 1968).

Gardin, J.C., Studies on the Automatic Indexing of Scientific Documents [in French], (Cntr. Nat'l. Rech. Sci.); Rev. Fran. Info. Et Rech. Op. 1, No. 6, 27-46 (1967).

Gardin, J.C., SYNTOL, Vol. II., of the Rutgers Series on Systems for the Intellectual Organization of Information, Ed. S. Artandi, 106 p. (Graduate School of Library Science, Rutgers, The State Univ., New Brunswick, N.J., 1965).

Garfield, E., Can Citation Indexing Be Automated?, in Statistical Association Methods For Mechanized Documentation, Symp. Proc., Washington, D.C., Mar. 17-19, 1964, NBS Misc. Pub. 269, Ed. M.E. Stevens et al, pp. 189-192 (U.S. Govt. Print. Off., Washington, D.C., Dec. 15, 1965).

Garfield, E., Patent Citation Indexing and the Notions of Novelty, Similarity, and Relevance, J. Chem. Doc. 6, No. 2, 63-65 (May 1966).

Garfield, E., Science Citation Index --- Answers to Frequently Asked Questions, Rev. Int. Doc. 32, 112-116 (Aug. 1965).

Garfield, E., I. H. Sher and R. J. Torpie, The Use of Citation Data in Writing the History of Science, 86 p. (Institute for Scientific Information, Philadelphia, Pa., 1964).

Garvin, P. L., Automatic Linguistic Analysis --- A Heuristic Problem, Proc. Int. Conf. on Machine Translation of Languages and Applied Language Analysis, Symp. 13, Vol. II, National Physical Laboratory, Teddington, England, Sept. 5-8, 1961, pp. 656-671 (Her Majesty's Stationery Office, London, 1962).

Garvin, P. L., Heuristic Syntax. Computer-Aided Research in Machine Translation, Progress Rept. 14, 36 p. (Bunker-Ramo Corp., Canoga Park, Calif., Mar. 27, 1967).

Garvin, P. L., An Informal Survey of Modern Linguistics, Am. Doc. 16, No. 4, 291-298 (Oct. 1965).

Garvin, P. L., Language and Machines, Int. Sci. & Tech., No. 65, 63-76 (May 1967).

Garvin, P. L., Machine Translation --- Fact or Fancy?, Datamation 13, No. 4, 29-31 (April 1967).

Garvin, P. L., Ed., Natural Language and the Computer, 398 p. (McGraw-Hill, New York, 1963).

Garvin, P. L., Some Comments on Algorithm and Grammar in the Automatic Parsing of Natural Languages, 10 p. (Bunker-Ramo Corp., Canoga Park, Calif., 1965).

Garvin, P. L., Some Comments on Algorithm and Grammar in the Automatic Parsing of Natural Languages, Mech. Trans. 9, No. 1, 2-3 (Mar. 1966).

Garvin, P. L., A Syntactic Analyzer Study, Final Rept. No. RADC-TR-65-309 on Contract AF30(602)-3506, 187 p. (Bunker-Ramo Corp., Canoga Park, Calif., 1965).

Garvin, P. L. and B. Spolsky, Eds., Computation in Linguistics: A Case Book, 340 p. (Indiana Univ. Press, Bloomington, 1966).

Garvin, P. L., J. Brewer and M. Mathiot, Prediction-Typing --- A Pilot Study in Semantic Analysis, 88 p. (Bunker-Ramo Corp., Canoga Park, Calif., Jan. 1966).

Geddes, E. W., R. L. Emrich and J. F. McMurrer, Feasibility Report and Recommendations for New York State Identification and Intelligence System, Tech. Memo. No. TM-LO-1000/000/00, 276 p. (System Development Corp., Santa Monica, Calif., Nov. 1, 1963).

General Electric Company, Technical Military Planning Operation (TEMPO), Information System Research, The DEACON Project, Research Memo. No. RM-65-TMP-69, Final Rept. for the period April 1, 1964 - Sept. 30, 1965, 48 p. (General Electric Co., Oct. 1965).

General Electric Company, Technical Military Planning Operation (TEMPO), Phase-Structure Oriented Targeting Query Language, The DEACON Project, Research Memo. No. RM-65-TMP-64, Final Rept, 18 p. (General Electric Co., Sept. 1965).

George, A. L., Qualitative and Quantitative Procedures in Content Analysis, 38 p. (The RAND Corp., Santa Monica, Calif., Dec. 1964).

Ghizzetti, A., Ed., Automatic Translation of Languages, papers presented at the NATO Summer School, Venice, 1962, 242 p. (Pergamon, New York, 1967).

Gifford, C. and G.J. Baumanis, On Understanding User Choices: Textual Correlates of Relevance Judgments, Am. Doc. 20, No. 1, 21-26 (Jan. 1969).

Ginsberg, H.F., R.F. Schmitz, W.K. Holman and M.D. Hall, Computer Aids in the Evaluation of Indexing Terminology, J. Chem. Doc. 7, No. 4, 237-239 (Nov. 1967).

Giuliano, V.E., The Interpretation of Word Associations, in Statistical Association Methods For Mechanized Documentation, Symp. Proc., Washington, D.C., Mar. 17-19, 1964, NBS Misc. Pub. 269, Ed. M.E. Stevens et al, pp. 25-32 (U.S. Govt. Print. Off., Washington, D.C., Dec. 15, 1965).

Giuliano, V.E., Postscript: A Personal Reaction to Reading the Conference Manuscripts, in Statistical Association Methods For Mechanized Documentation, Symp. Proc., Washington, D.C., Mar. 17-19, 1964, NBS Misc. Pub. 269, Ed. M.E. Stevens et al, pp. 259-260 (U.S. Govt. Print. Off., Washington, D.C., Dec. 15, 1965).

Giuliano, V.E. and P.E. Jones, Study and Test of a Methodology for Laboratory Evaluation of Message Retrieval Systems, 191 p. (U.S. Air Force, Bedford, Mass., Aug. 1966).

Gold, B., Word-Recognition Computer Program, Rept. No. TR-452, 27 p. (Research Lab., of Electronics, M.I.T., Cambridge, Mass., June 15, 1966).

Goldhor, H., Ed., Proc. 1963 Clinic on Library Applications of Data Processing, Univ. of Illinois, Apr. 28 - May 1, 1963, 176 p. (Graduate School of Library Science, Illinois Univ., Urbana, 1964).

Gor'Kova, V.I. and K.I. Naumycheva, Preparation of Titles of Secondary Scientific Documents, Auto. Documentation & Math. Linguistics 1, No. 1, 1-3 (Spring 1967).

Gotlieb, C.C. and S. Kumar, Semantic Clustering of Index Terms, J. ACM 15, No. 4, 493-513 (Oct. 1968).

Gould, L. and S.M. Lamb, Type-Lists, Indexes, and Concordances from Computers, 112 p. (Yale Univ., Linguistic Automation Project, New Haven, Conn., Feb. 1967).

Granito, C.E., A. Gelberg, J.E. Schultz, G.W. Gibson and E.A. Metcalf, Rapid Structure Searches via Permuted Chemical Line Notations. II. A Key-Punch Procedure for the Generation of an Index for a Small File, J. Chem. Doc. 5, No. 1, 52-55 (Feb. 1965).

Granito, C.E., J.E. Schultz, G.W. Gibson, A. Gelberg, R.J. Williams and E.A. Metcalf, Rapid Structure Searches via Permuted Chemical Line Notations. III. A Computer-Produced Index, J. Chem. Doc. 5, No. 4, 229-233 (Nov. 1965).

Grauer, R.T. and M. Messier, An Evaluation of Rocchio's Clustering Algorithm, in Information Storage and Retrieval, Rept. No. ISR-12, [G. Salton, Proj. Dir.] pp. VI-1 to VI-39 (Cornell Univ., Ithaca, N.Y., June 1967).

Graves, P.A., D.G. Hays, M. Kay and T.W. Ziehe, Computer Routines to Read Natural Text with Complex Formats, Research Memo. No. RM-4920-PR, 141 p. (The RAND Corp., Santa Monica, Calif., Aug. 1966).

Green, C.C. and B. Raphael, The Use of Theorem-Proving Techniques in Question-Answering Systems, Proc. 23rd National Conf., ACM, Las Vegas, Nev., Aug. 27-29, 1968, pp. 169-181 (Brandon/Systems Press, Inc., Princeton, N.J., 1968).

Greene, M., A Reference-Connecting Technique for Automatic Information Classification and Retrieval, Rept. No. OEG Research contrib. -77, 22 p. (Center for Naval Analyses, Operations Evaluation Group, Washington, D.C., Mar. 10, 1967).

Grenander, U., A Feature Logic for Clusters, 29 p. (IBM Corp., Cambridge Research Center, Cambridge, Mass., June 1968).

Grenander, U., Syntax-Controlled Probabilities, 29 p. (Division of Applied Mathematics, Brown Univ., Providence, R.I., July 1969).

Grover, N.B., B. Arbel, J. Gross and C. Keran, Computer Retrieval of Mechanically-Indexed Articles, Methods of Inf. in Medicine 7, No. 4, 224-226 (Oct. 1968).

Gull, C.D., Letter to the Editor, Am. Doc. 18, No. 4, 252-253 (Oct. 1967).

Hammond, W., Statistical Association Methods for Simultaneous Searching of Multiple Document Collections, in Statistical Association Methods For Mechanized Documentation, Symp. Proc., Washington, D.C., Mar. 17-19, 1964, NBS Misc. Pub. 269, Ed. M.E. Stevens et al, pp. 237-239 (U.S. Govt. Print. Off., Washington, D.C., Dec. 15, 1965).

Hammond, W.B., Vocabulary Construction and Control, Proc. Workshop on Working with Semi-Automatic Documentation Systems, Warrenton, Va., May 2-5, 1965, Ed. J.J. Maher, pp. 71-73 (System Development Corp., Santa Monica, Calif., 1965).

Harper, K.E., Contextual Analysis, Mech. Trans. 4, No. 3, 70-75 (Dec. 1957).

Harper, K.E., Contextual Analysis in Word-for-Word MT, Mech. Trans. 3, No. 2, 40-41 (Nov. 1956).

Harper, K.E., Studies in Inter-Sentence Connection, Research Memo. No. RM-4828-PR, 21 p. (The RAND Corp., Santa Monica, Calif., Dec. 1965).

Harper, K.E. and S.Y.W. Su, A Directed Random Paragraph Generator, Research Memo. No. RM-6053-PR, 13 p. (The RAND Corp., Santa Monica, Calif., July 1969).

Harper, K.E., D.G. Hays, D.S. Worth and T.W. Ziehe, Six Tasks in Computational Linguistics, Research Memo. No. RM-2803-AFOSR, 107 p. (The RAND Corp., Santa Monica, Calif., Oct. 1961).

Harris, D.J. and A. Kent, The Computer as an Aid to Lawyers, The Computer J. 10, No. 1, 22-28 (May 1967).

Harris, Z.S., Discourse Analysis, Language 28, 1-30 (1952).

Harris, Z.S., String Analysis of Sentence Structure, 70 p. (Mouton, The Hague, Netherlands, 1962).

Harris, Z.S., Transformational Theory, Language 41, No. 1, 363-401 (July-Sept. 1965).

Harvard University Computation Lab., Mathematic Linguistics and Automatic Translation, Rept. No. NSF-17 (Cambridge, Mass., Aug. 1966).

Hayes, R.M., The Decomposition of Vocabulary Hierarchies, in Mechanized Information Storage, Retrieval and Dissemination, Proc. F.I.D./I.F.I.P. Joint Conf., Rome, Italy, June 14-17, 1967, Ed. K. Samuelson, pp. 160-191 (North-Holland Pub. Co., Amsterdam, 1968).

Hays, D.G., Automatic Content Analysis: Some Entries for a Transformation Catalog, 21 p. (The RAND Corp., Santa Monica, Calif., April 1960).

Hays, D.G., Automatic Language-Data Processing in Sociology, 33 p. (The RAND Corp., Santa Monica, Calif., Sept. 1959).

Hays, D.G., Computational Linguistics: Research in Progress at The RAND Corporation, 10 p. (The RAND Corp., Santa Monica, Calif., Aug. 1966).

Hays, D.G., An Essay on the Profession of Linguistics as a Customer for Automatic Documentation, in Information in the Language Sciences, Proc. Conf. on Information in the Language Sciences, Warrenton, Va., Mar. 4-6, 1966, sponsored by the Center for Applied Linguistics, Ed. R.R. Freeman et al, pp. 57-65 (American Elsevier Pub. Co., New York, 1968).

Hays, D.G., Introduction to Computational Linguistics, 260 p. (American Elsevier Pub. Co., New York, 1967).

Hays, D.G., Linguistic Foundations for a Theory of Content Analysis, 21 p. (The RAND Corp., Santa Monica, Calif., Nov. 1967).

Hays, D.G., Parsing, in Readings in Automatic Language Processing, Ed. D.G. Hays, pp. 73-82 (American Elsevier Pub. Co., New York, 1966).

Hays, D.G., Processing Natural Language Text, 8 p. (The RAND Corp., Santa Monica, Calif., Oct. 1966).

Hays, D.G., Ed., Readings in Automatic Language Processing, 202 p. (American Elsevier Pub. Co., New York, 1966).

Hays, D.G., Report of a Summer Seminar on Computational Linguistics, Research Memo. No. RM-3889-NSF, 57 p. (The RAND Corp., Santa Monica, Calif., Feb. 1964).

Hays, D.G., Research Procedures in Machine Translation, in Natural Language and the Computer, Ed. P.L. Garvin, pp. 183-214 (McGraw-Hill, New York, 1963).

Hays, D.G. and M.L. Rapp, Annotated Bibliography of RAND Publications in Computational Linguistics, Research Memo. No. RM-3894-2, 29 p. (The RAND Corp., Santa Monica, Calif., April 1966).

Hays, D.G., B. Henisz-Dostert and M. Rapp, Computational Linguistics: Bibliography, 1965, Research Memo. No. RM-4986-PR, 80 p. (The RAND Corp., Santa Monica, Calif., April 1966).

Hays, D.G., B. Henisz-Dostert and M. Rapp., Computational Linguistics: Bibliography, 1966, Research Memo. No. RM-5345-PR, 90 p. (The RAND Corp., Santa Monica, Calif., April 1967).

Heald, J.H., Transition from a Manual to a Machine Indexing System, in Machine Indexing: Progress and Problems, Proc. Third Institute on Information Storage and Retrieval, Washington, D.C., Feb. 13-17, 1961, pp. 170-190 (The American Univ., Washington, D.C., 1962).

Heiser, R.C. and D.J. Hillman, A Formal Theory of Conceptual Affiliation for Document Reconstruction, Report No. 1: The Affiliation-Value Model, 30 p. (Center for the Information Sciences, Lehigh Univ., Bethlehem, Pa., 1966).

Herbert, E., Information Transfer, Int. Sci. & Tech. No. 51, 26-37 (Mar. 1966).

Hersey, D.F. and W. Hammond, Computer Usage in the Development of a Water Resources Thesaurus, Am. Doc. 18, No. 4, 209-215 (Oct. 1967).

Heward, F.C., Information Indexing by Computer, Post Office Elec. J. 60, Part I, 28-32 (April 1967).

Hill, D.R., A Vector Cluster Technique, in Mechanized Information Storage, Retrieval and Dissemination, Proc. F.I.D./I.F.I.P. Joint Conf., Rome, Italy, June 14-17, 1967, Ed. K. Samuelson, pp. 225-234 (North-Holland Pub. Co., Amsterdam, 1968).

Hillman, D.J., An Algorithm for Document Characterization, Rept. No. 2, Mathematical Theories of Relevance with Respect to the Problems of Indexing, 56 p. (Center for the Information Sciences, Lehigh Univ., Bethlehem, Pa., Mar. 12, 1965).

Hillman, D.J., Document Retrieval Theory, Relevance, and the Methodology of Evaluation, Report No. 1: Characterization and Connectivity, 41 p. (Center for the Information Sciences, Lehigh Univ., Bethlehem, Pa., May 24, 1966).

Hillman, D.J., Document Retrieval Theory, Relevance, and the Methodology of Evaluation, Report No. 5: Arithmatization of Syntactic Analysis, 19 p. (Center for the Information Sciences, Lehigh Univ., Bethlehem, Pa., July 29, 1967).

Hillman, D.J., Grammars and Text Analysis, Computational, Phonological and Morphological Linguistics and Retrieval Studies, Report No. 1, 20 p. (Center for the Information Sciences, Lehigh Univ., Bethlehem, Pa., Aug. 1965).

Hillman, D.J., Mathematical Classification Techniques for Non-Static Document Collections, with Particular Reference to the Problem of Relevance, in Proc. Second Study Conf. on Classification Research, Elsinore, Denmark, Sept. 14-18, 1964, pp. 177-209 (Munksgaard, Copenhagen, 1965).

Hillman, D.J., Negotiation of Inquiries in an On-Line Retrieval System, Inf. Storage & Retrieval 4, No. 2, 219-238 (June 1968).

Hillman, D.J. and A.J. Kasarda, The LEADER Retrieval System, AFIPS Proc. Spring Joint Computer Conf., Vol. 34, Boston, Mass., May 14-16, 1969, pp. 447-455 (AFIPS Press, Montvale, N.J., 1969).

Himmelman, D.S. and J.T. Chu, An Automatic Abstracting Program Employing Stylo-Statistical Techniques and Hierarchical Data Indexing, in preprints of papers presented at the 16th National Meeting, ACM, Los Angeles, Calif., Sept. 5-8, 1961, pp. 5C-3(1) to 5C-3(3) (Assoc. for Computing Machinery, New York, 1961).

Hirayama, K. and K. Kango, Comparison of Keyword Indexing and Indexing by Systematic Classification, paper presented at the Congress of International Federation for Documentation (FID), Washington, D.C., Oct. 10-15, 1965, 14 p. (preprint).

Hochgesang, G.T., SOCCER --- A Concordance Program, in Information Storage and Retrieval, Rept. No. ISR-11, [G. Salton, Proj. Dir.] pp. III-1 to III-23 (Cornell Univ., Ithaca, N.Y., June 1966).

Holland, W.B., Relational Data File: Input to an Experimental File on Soviet Cybernetics, Research Memo. No. RM-5622-PR, 195 p. (The RAND Corp., Santa Monica, Calif., Feb. 1969).

Holst, W., Mechanical Translation by Coordinate Indexing, Am. Doc. 17, No. 3, 140-141 (July 1966).

Holsti, O.R., An Adaptation of the 'General Inquirer' for the Systematic Analysis of Political Documents, Behav. Sci. 9, 382-388 (1964).

Holsti, O.R., Computer Content Analysis in International Relations Research, in Computers in Humanistic Research, Ed. E.A. Bowles, pp. 108-117 (Prentice-Hall, Englewood Cliffs, N.J., 1967).

Holsti, O.R., A System of Automated Content Analysis of Documents (Stanford Univ., Palo Alto, Calif., Mimeo rev. ed., 1963).

Holsti, O.R., J.K. Loomba and R.C. North, Content Analysis, in The Handbook of Social Psychology, 2nd ed., Vol. 2, Research Methods, Ed. G. Lindzey and E. Aronson, pp. 596-692 (Addison Wesley Pub. Co., Reading, Mass., 1968).

Hooper, R.S., Indexer Consistency Tests --- Origin, Measurements, Results and Utilization, paper presented at the Congress of International Federation for Documentation (FID), Washington, D.C., Oct. 10-15, 1965, 19 p. (preprint).

Hormann, A., GAKU: An Artificial Student, Behav. Sci. 10, 88-107 (Jan. 1965).

Hormann, A.M., Introduction to ROVER, an Information Processor, Field Note No. FN-3487, 51 p. (System Development Corp., Santa Monica, Calif., Apr. 25, 1960).

Hormann, A., A New Task Environment for Gaku Teamed with a Man, Tech. Memo. No. TM-2311/003/00, 26 p. (System Development Corp., Santa Monica, Calif., May 27, 1966).

Horty, J.F., A Look at Research in Legal Information Retrieval, in Proc. Second International Study Conf. on Classification Research, Elsinore, Denmark, Sept. 14-18, 1964, Ed. P. Atherton, pp. 382-393 (Discussion, pp. 394-396), (Munksgaard, Copenhagen, 1965).

Horvath, P.J., A.Y. Chamis, R.F. Carroll and J. Dlugos, The B F Goodrich Information Retrieval System and Automatic Information Distribution Using Computer-Compiled Thesaurus and Dual Dictionary. J. Chem. Doc. 7, No. 3, 124-130 (Aug. 1967).

Houkes, J.M., Ed., Management Information Systems and the Information Specialist, Proc. Symp. Purdue Univ., July 12-13, 1965, 138 p. (Krannert Graduate School of Industrial Admin. and the University Libraries, Purdue Univ., Lafayette, Ind., 1966).

Householder, F.W. and J.P. Thorne, Automatic Language Analysis, 7th Quart. Rept., 38 p. (Indiana Univ., Bloomington, 1962).

Hunt, E.B., J. Maran and P.J. Stone, Experiments in Induction, 247 p. (Academic Press, New York, 1966).

Hurwitz, F.I., A Study of Indexer Consistency, Am. Doc. 20, No. 1, 92-94 (Jan. 1969).

Hutchins, W.J., Automatic Document Selection Without Indexing, J. Doc. 23, No. 4, 273-290 (Dec. 1967).

Hutchins, W.J., Automatic Document Selection, [Letter to the Editor], J. Doc. 24, No. 2, 119-120 (June 1968).

Hymes, D., Ed., The Use of Computers in Anthropology, Symp. on the Use of Computers in Anthropology, Wartenstein Castle, Austria, June 20-30, 1962, 558 p. (Mouton, The Hague, 1965).

Hyslop, M., Subject Indexing Using a Computer-Manipulated Thesaurus as Vocabulary Control, FID Congress, Oct. 10-15, 1965, 10 p. (preprint).

Iakushin, B.V., Problems of Algorithmic Composition of Subject Indexes, Rept. No. LT-65-102, 18 p. (Translation of foreign literature, Nauchno-Tekhnicheskaya Informatsiya (USSR) No. 5, 1965, pp. 22-25) (The RAND Corp., Santa Monica, Calif., Mar. 15, 1966).

Ide, E., R. Williamson and D. Williamson, The Cornell Programs for Cluster Searching and Relevance Feedback, in Information Storage and Retrieval, Rept. No. ISR-12 [G. Salton, Proj. Dir.] pp. IV-1 to IV-13 (Cornell Univ., Ithaca, N.Y., June 1967).

Iker, H.P. and N.I. Harway, A Computer Approach Towards the Analysis of Content, Behav. Sci. 10, No. 2, 173-182 (1965).

Index de la Littérature Nucléaire Française, Vol. I, No. 1, 1968 (Centre d'Etudes Nucléaires, Saclay, 1968).

International Conference on Computational Linguistics, New York, N.Y., May 19-21, 1965, Preprints, Association for Machine Translation and Computational Linguistics, 1965, 1 vol. (Loose-leaf).

Jackson, D.M., The Construction of Retrieval Environments and Pseudo-Classifications Based on External Relevance, Computer and Information Science Tech. Rept. No. 69-3, 74 p. (The Ohio State Univ., Columbus, April 1969).

Jacobson, S.N., A Modified Routine for Connecting Related Sentences of English Text, in Computation in Linguistics: A Case Book, Ed. P.L. Garvin and B. Spolsky, pp. 284-311 (Indiana Univ. Press, Bloomington, 1966).

Jaffe, J., The Study of Language in Psychiatry Psycholinguistics and Computational Linguistics, reprinted from American Handbook of Psychiatry, Vol. 3, Ed. A. Silvano, pp. 689-704 (Basic Books, Inc., 1959).

Jaffe, J., Verbal Behavioral Analysis in Psychiatric Interviews with the Aid of Digital Computers, in Disorders of Communication: Res. Publ. Assoc. Res. Nerv. Ment. Dis., Vol. 42, Ed. D.McK. Rioch and E.A. Weinstein, Chapter 27, 12 p. (Williams and Wilkins, Baltimore, Md., 1964).

Jahoda, G., J.J. Oliva and A.J. Dean, Recall with Keyword from Title Indexes: Effect of Question-Relevant Document Title Concept Correspondence, 1 v. (Library School, Florida State Univ., Tallahassee, Dec. 1966).

Jahoda, G. and M.L. Stursa, Test of Indexes. A Comparison of Keyword from Title Indexes With and Without Added Keywords and a Single Access Point per Document Alphabetic Subject Index, 55 p. (School of Library Science, Florida State Univ., Tallahassee, Jan. 1969).

Janda, K., Computer Applications in Political Science, AFIPS Proc. Fall Joint Computer Conf., Vol. 31, Anaheim, Calif., Nov. 14-16, 1967, pp. 339-345 (Thompson Books, Washington, D.C., 1967).

Jernigan, R. and A.G. Dale, Set Theoretic Models for Classification and Retrieval, Rept. No. LRC-64-WTM-5, 20 p. (Linguistics Research Center, Texas Univ., Austin, Nov. 1964).

Johanningsmeier, W.F. and F.W. Lancaster, Project SHARP (SHips Analysis and Retrieval Project) Information Storage and Retrieval System: Evaluation of Indexing Procedures and Retrieval Effectiveness, Rept. No. NAVSHIPS 250-210-3, 49 p. (Department of the Navy, Washington, D.C., June 1964).

Johnpoll, B.K., The Canada News Index: A Report on Computerized Indexing of News in Selected Canadian Dailies, Spec. Lib. 58, No. 2, 102-105 (Feb. 1967).

Johnson, D.L. and R.E. Wall, Logical Processing and Context Analysis in Mechanical Translation, Trend in Engn. at the University of Washington 10, No. 3, 14-22 (July 1958).

Jones, P.E., Historical Foundations of Research on Statistical Association Techniques for Mechanized Documentation, in Statistical Association Methods For Mechanized Documentation, Symp. Proc., Washington, D.C., Mar. 17-19, 1964, NBS Misc. Pub. 269, Ed. M.E. Stevens et al, pp. 3-8 (U.S. Govt. Print. Off., Washington, D.C., Dec. 15, 1965).

Jones, P.E. and R.M. Curtice, A Framework for Comparing Term Association Measures, Am. Doc. 18, No. 3, 153-161 (July 1967).

Jones, P.E., V.E. Giuliano and R.M. Curtice, Linear Models for Associative Retrieval, Rept. No. ESD-TR-67-202, 215 p., Vol. II (Decision Sciences Lab., Electronic Systems Div., A.F. Systems Command, L.G. Hanscom Field, Bedford, Mass., Feb. 1967).

Jones, P.E., V.E. Giuliano and R.M. Curtice, Papers on Automatic Language Processing, Vol. I, Selected Collection Statistics and Data Analysis, 122 p. (Arthur D. Little, Inc., Cambridge, Mass., Feb. 1967).

Jones, P.E., V.E. Giuliano and R.M. Curtice, Papers on Automatic Language Processing, Vol. III, Development of String Indexing Techniques, 142 p. (Arthur D. Little, Inc., Cambridge, Mass., Feb. 1967).

Jordan, J.R. and W.J. Watkins, KWOC Index as an Automatic By-Product of SDI, in Information Transfer, Proc. Am. Soc. Inf. Sci. Annual Meeting, Vol. 5, Columbus, O., Oct. 20-24, 1968, pp. 211-215 (Greenwood Pub. Corp., New York, 1968).

Joynes, M.L., Automatic Verification of Phrase Structure Description, in Computation in Linguistics: A Case Book, Ed. P.L. Garvin and B. Spolsky, pp. 183-206 (Indiana Univ. Press, Bloomington, 1966).

Juhasz, S., Ed., MAMMAX (Machine Made and Machine Aided Indexes), in National Federation of Science Abstracting and Indexing Services, Proc. Annual Meeting, Philadelphia, Pa., March 1967, 143 p.

Juhasz, S., H. Wooster, E.A. Ripperger and D. Falconer, Extended WADEX System: Tool for Browsing, Searching, and Express Information with Adjustable Intellectual Preparation Effort, 20 p. [paper presented at F.I.D. Congress, Washington, D.C., Oct. 12, 1965], (Applied Mechanics Reviews, San Antonio, Texas).

Jung, E., Permutierte Register mit Hilfe Elektronischer Datenverarbeitungsanlagen (Eine Einfuhrung). (Computer-Aided Permutation Indexes, An Introduction), ZIID-Zeitschrift 14, No. 6, 187-190 (Dec. 1967).

Kalenich, W.A., Ed., Information Processing 1965, Proc. IFIP Congress 65, Vol. 1, New York, N.Y., May 24-29, 1965, 304 p. (Spartan Books, Washington, D.C., 1965).

Kalenich, W.A., Ed., Information Processing 1965, Proc. IFIP Congress 65, Vol. 2, New York, N.Y., May 24-29, 1965, pp. 305-648 (Spartan Books, Washington, D.C., 1966).

Kaplan, N., The Norms of Citation Behavior: Prolegomena to the Footnote, Am. Doc. 16, No. 3, 179-184 (July 1965).

Katz, J.H., Simulation of Outpatient Appointment Systems, Commun. ACM 12, No. 4, 215-222 (April 1969).

Katz, J.J. and J.A. Fodor, The Structure of a Semantic Theory, Language 39, 170-210 (1963). Reprinted in part in, The Structure of Language; Readings in the Philosophy of Language, by J.A. Fodor and J.J. Katz (Prentice-Hall, Englewood Cliffs, N.J., 1964).

Kaufman, S., The IBM Information Retrieval Center - (ITIRC) System Techniques and Applications, Proc. 21st National Conf., ACM, Los Angeles, Calif., Aug. 30 - Sept. 1, 1966, pp. 505-512 (Thompson Book Co., Washington, D.C., 1966).

Kay, M., Experiments with a Powerful Parser, Research Memo. No. RM-5452-PR, 33 p. (The RAND Corp., Santa Monica, Calif., Oct. 1967). Also in 2 ème Conf. Internationale sur le Traitement Automâtique des Langues, Grenoble, Aug. 1967.

Kay, M., From Semantics to Syntax, Rept. No. P-3746, 28 p. (The RAND Corp., Santa Monica, Calif., Dec. 1967).

Kay, M., The Tabular Parser: A Parsing Program for Phrase Structure and Dependency, Research Memo. No. RM-4933-PR, 49 p. (The RAND Corp., Santa Monica, Calif., July 1966).

Kay, M. and T.W. Ziehe, Natural Language in Computer Form, in Readings in Automatic Language Processing, Ed. D.G. Hays, pp. 33-49 (American Elsevier Pub. Co., New York, 1966).

Kay, M. and T.W. Ziehe, Natural Language in Computer Form, Research Memo. No. RM-4390, 81 p. (The RAND Corp., Santa Monica, Calif., Feb. 1965).

Keen, E.M., Document Length, in Information Storage and Retrieval, Rept. No. ISR-13, [G. Salton, Proj. Dir.] pp. V-1 to V-60 (Cornell Univ., Ithaca, N.Y., Jan. 1968).

Keen, E.M., Search Matching Functions, in Information Storage and Retrieval, Rept. No. ISR-13 [G. Salton, Proj. Dir.] pp. III-1 to III-58 (Cornell Univ., Ithaca, N.Y., Jan. 1968).

Keen, E.M., Suffix Dictionaries, in Information Storage and Retrieval, Rept. No. ISR-13 [G. Salton, Proj. Dir.] pp. VI-1 to VI-22 (Cornell Univ., Ithaca, N.Y., Jan. 1968).

Keen, E.M., Thesaurus, Phrase and Hierarchy Dictionaries, in Information Storage and Retrieval, Rept. No. ISR-13 [G. Salton, Proj. Dir.] pp. VII-1 to VII-59 (Cornell Univ., Ithaca, N.Y., Jan. 1968).

Kehl, W.B., Computers and Literature, Data Proc. Mag. 7, 24-26 (July 1965).

Kellogg, C.H., CONVERSE --- A System for the On-Line Description and Retrieval of Structural Data Using Natural Language, Rept. No. SP-2635, 16 p. (System Development Corp., Santa Monica, Calif., May 26, 1967). Also in Mechanized Information Storage, Retrieval and Dissemination, Proc. F.I.D./I.F.I.P. Joint Conf., Rome, Italy, June 14-17, 1967, Ed. K. Samuelson, pp. 608-621 (North-Holland Pub. Co., Amsterdam, 1968).

Kellogg, C.H., On-Line Translation of Natural Language Questions into Artificial Language Queries, Inf. Storage & Retrieval 4, No. 3, 287-307 (Aug. 1968).

Kent, A., Textbook on Mechanized Information Retrieval, 268 p. (Interscience Pub., New York, 1962); 2nd ed., 371 p. (Interscience Pub., N.Y., 1966).

Kent, A., O.E. Taulbee, J. Belzer and G.D. Goldstein, Eds., Electronic Handling of Information: Testing and Evaluation, 313 p. (Thompson Book Co., Washington, D.C., 1967).

Kent, A.K., Comparison of the Philosophy of Indexing Text with that of Indexing Structural Formulae, Aslib Proc. 19, No. 11, 364-368 (Nov. 1967).

Kessler, M.M., Comparison of the Results of Bibliographic Coupling and Analytic Subject Indexing, Am. Doc. 16, No. 3, 223-233 (July 1965).

Kessler, M.M., Some Statistical Properties of Citations in the Literature of Physics, in Statistical Association Methods For Mechanized Documentation, Symp. Proc., Washington, D.C., Mar. 17-19, 1964, NBS Misc. Pub. 269, Ed. M.E. Stevens et al, pp. 193-198 (U. S. Govt. Print. Off., Washington, D.C., Dec. 15, 1965).

Keyser, S.J. and R. Kirk, Machine Recognition of Transformational Grammars of English, 74 p. (Brandeis Univ., Waltham, Mass., 1967).

Kikuchi, T., Automatic Indexing and Automatic Classification: A Review. I, [in Japanese] Jo ho Kanri, JICST 8, No. 13, 3-9 (1966).

Kirsch, R.A., Computer Interpretation of English Text and Picture Patterns, IEEE Trans. Electron. Computers EC-13, No. 4, 363-376 (Aug. 1964).

Klein, S., Automatic Paraphrasing in Essay Format, Mech. Trans. 8, No. 3-4, 68-83 (June-Oct. 1965).

Klein, S. and R.F. Simmons, A Computational Approach to Grammatical Coding of English Words, J. ACM 10, No. 3, 334-347 (July 1963).

Klein, S. and R. F. Simmons, Syntactic Dependence and the Computer Generation of Coherent Discourse, Mech. Trans. 7, No. 2, 50-61 (Aug. 1963).

Klein, S., B.G. Davis, W. Fabens, R.G. Herriot, B.J. Katke, M.A. Kuppin and A.E. Towster, AUTOLING: An Automated Linguistic Field-Worker, in Second Conference Internationale sur le Traitement Automatique des Langues, Grenoble, Aug. 1967, Paper No. 31.

Knowlton, K.C., Sentence Parsing with a Self-Organizing Heuristic Program, Ph.D. Thesis, 1 v. (Research Lab. of Electronics, M.I.T., Cambridge, Mass., Sept. 1962).

Koch, K.H., Internationale Dezimal Klassifikation (DK) und Elektronische Datenverarbetung, 67 p. (Zentralstelle für Maschinelle Dokumentation, Frankfurt/Main, Dec. 1967).

Kochen, M., Ed., The Growth of Knowledge (Readings on Organization and Retrieval of Information), 394 p. (Wiley, New York, 1967).

Kochen, M., Introduction, in Some Problems in Information Science, Ed. M. Kochen, pp. 11-60 (The Scarecrow Press, Inc., New York, 1965).

Kochen, M., The Knowledge Subsystem, in Some Problems in Information Science, Ed. M. Kochen, pp. 61-66 (The Scarecrow Press, Inc., New York, 1965).

Kochen, M., Ed., Some Problems in Information Science, 309 p. (The Scarecrow Press, Inc., New York, 1965).

Kochen, M., Stability in the Growth of Knowledge, Am. Doc. 20, No. 3, 186-197 (July 1969).

Kochen, M., The Storage/Recall Subsystem, in Some Problems in Information Science, Ed. M. Kochen, pp. 151-155 (The Scarecrow Press, Inc., New York, 1965).

Kochen, M. and L. Uhr, A Model for the Process of Learning to Comprehend, in Some Problems in Information Science, Ed. M. Kochen, pp. 94-104 (The Scarecrow Press, Inc., New York, 1965).

Kochen, M. and R. Tagliacozzo, Book-Indexes as Building Blocks for a Cumulative Index, Am. Doc. 18, No. 2, 59-66 (April 1967).

Kollin, R. and J.L. Harris, A Criterion for Evaluation of Indexing Systems, in Information Transfer, Proc. Am. Soc. Inf. Sci. Annual Meeting, Vol. 5, Columbus, O., Oct. 20-24, 1968, pp. 79-81 (Greenwood Pub. Corp., New York, 1968).

Koltovoj, B., Literature Automation, in Izvestiya, Jan. 5, 1967, p. 4, as translated and summarized by P. Stephan, in Soviet Cybernetics: Recent News Items No. 2, Ed. W.B. Holland, pp. 38-41 (The RAND Corp., Santa Monica, Calif., Mar. 1967).

Korotkin, A., L.H. Oliver and D.R. Burgis, Indexing Aids, Procedures and Devices, Rept. No. RADC-TR-64-582, 110 p. (General Electric Co., Bethesda, Md., April 1965).

Korvasova, K., Mechanical Analysis of Source Language, Inf. Proc. Machines, No. 12, 99-106 (1966).

Köster, K., The Use of Computers in Compiling National Bibliographies: Illustrated by the Example of the Deutsche Bibliographie, Libri 16, No. 4, 269-281 (1966).

Kozumplik, W.A. and R.T. Lange, Computer-Produced Microfilm Library Catalog, Am. Doc. 18, No. 2, 67-80 (April 1967).

Krallman, D., Statistische Methoden in der Stilistischen Textanalyse: Ein Beitrag zur Informationserschliefzung Mithilfe Elektronischer Rechenmaschinen, Doctoral Dissertation, Philosophischein Fakultät der Rhenischen Friedrich-Wilhelms-Universität zn Bonn, Bonn, 1966, 43 p.

Kreithen, A., Vocabulary Control in Automatic Indexing, Data Proc. Mag. 7, No. 2, 60-61 (1965).

Krollman, F., H.J. Schuck and U. Winkler, Production of Text-Related Technical Glossaries by Digital Computer; A Procedure to Provide an Automatic Translation Aid, 33 p. (Defense Language Institute, Washington, D.C., 1965). [Translated from the German].

Kučera, H. and W.N. Francis, Computational Analysis of Present-Day American English, 424 p. (Brown Univ. Press, Providence, R.I., 1967).

Kuhns, J.L., The Continuum of Coefficients of Association, in Statistical Association Methods For Mechanized Documentation, Symp. Proc., Washington, D.C., Mar. 17-19, 1964, NBS Misc. Pub. 269, Ed. M.E. Stevens et al, pp. 33-39 (U.S. Govt. Print. Off., Washington, D.C., Dec. 15, 1965).

Kulagina, O.S. and I.A. Mel'chuk, Automatic Translation: Some Theoretical Aspects and the Design of a Translation System, in Machine Translation, Ed. A.D. Booth, pp. 137-171 (Wiley, New York, 1967).

Kulkarni, J.M., Automatic Document Classification in Information Retrieval System, Herald Libr. Sci. 4, No. 2, 140-148 (April 1965).

Kuney, J.H., Publication and Distribution of Information, in Annual Review of Information Science and Technology, Vol. 3, Ed. C.A. Cuadra, pp. 31-59 (Encyclopedia Britannica, Inc., Chicago, Ill., 1968).

Kuno, S., Computer Analysis of Natural Languages, in Mathematical Aspects of Computer Science, Proc. Symposia in Applied Mathematics, Vol. 19, New York, N.Y., Apr. 5-7, 1966, Ed. J.T. Schwarts, pp. 52-110 (Amer. Mathematic Soc., Providence, R.I., 1967).

Kuno. S., Current Research in Computational and Mathematical Linguistics at the Computation Laboratory of Harvard University, in Mathematical Linguistics and Automatic Translation, Rept. No. NSF-15 to the National Science Foundation, Ed. A.G. Oettinger, pp. I-1 to I-9 (Computation Lab., Harvard Univ., Cambridge, Mass., Aug. 1965).

Kuno, S., Mathematic Linguistics and Automatic Translation, 263 p., Aug. 1967; 340 p. Sept. 1967 (Computation Lab., Harvard Univ., Cambridge, Mass.).

Kuno, S., The Predictive Analyzer and a Path Elimination Technique, Commun. ACM 8, No. 7, 453-462 (July 1965).

Kuno, S., A System for Transformational Analysis, in Mathematical Linguistics and Automatic Translation, Rept. No. NSF-15 to the National Science Foundation, Ed. A.G. Oettinger, pp. IV-1 to IV-44 (Computation Lab., Harvard Univ., Cambridge, Mass., Aug. 1965).

Kurfeerst, M. and J.W. Asher, A Factor Analysis of the Education Laws of Pennsylvania, Inf. Stor. & Retr. 4, No. 3, 257-270 (Aug. 1968).

Kurmey, W.J., An Evaluation of Automatically Prepared Abstracts and Indexes, M.S. Thesis, Chicago Univ., Ill., June 1964, 64 p.

Kuroda, S-Y., A Sketch of a Target Query Language, in Mathematic Linguistics and Automatic Translation, 13 p. (Computation Lab., Harvard Univ., Cambridge, Mass., Aug. 1967, Sec. 10).

Kyle, B.F., Information Retrieval and Subject Indexing: Cranfield and After, J. Doc. 20, No. 2, 55-69 (June 1964).

Laboratoire d'Automatique Documentaire de Linguistique, Centre National de la Recherche Scientifique (CNRS), "Etudes sur l'Indexation Automatique" [Studies on Automatic Indexing], Final Report, Contract 65-FR-160, 166 p. (Délégation Générale à la Recherche Scientifique et Technique, Marseilles, France, 1968).

Lamb, S.M., Linguistic Data Processing, in The Use of Computers in Anthropology, Symp. on the Use of Computers in Anthropology, Wartenstein Castle, Austria, June 20-30, 1962, Ed. D. Hymes, pp. 161-190 (Mouton, The Hague, 1965).

Lamb, S.M., On the Mechanization of Syntactic Analysis, in Readings in Automatic Language Processing, Ed. D.G. Hays, pp. 149-157 (American Elsevier Pub., New York, 1966).

Lamb, S.M. and L. Gould, Concordances From Computers, 90 p. (Univ. of California, Berkeley, 1964).

Lamson, B.G. and B. Dimsdale, A Natural Language Information Retrieval System, Proc. IEEE 54, 1636-1640 (Dec. 1966).

Lancaster, F.W., Information Retrieval Systems: Characteristics, Testing and Evaluation, 222 p. (Wiley, New York, 1968).

Lance, G.N. and W.T. Williams, Computer Programs for Hierarchical Polythetic Classification (Similarity Analyses), The Comput. J. 9, 60-64 (Mar. 1966).

Lance, G.N. and W.T. Williams, Computer Programs for Monothetic Classification ("Association Analysis"), The Comput. J. 8, No. 3, 246-249 (Oct. 1965).

Lance, G.N. and W.T. Williams, A General Theory of Classificatory Sorting Strategies II. Clustering Systems, The Comput. J. 10, No. 3, 271-277 (Nov. 1967).

Lazarow, A., E.H. Brekhus, K. Goodman and D.C. Norris, Computer Analysis of Document Content, in Information Retrieval with special reference to the Biomedical Sciences, papers presented at the Second Institute on Information Retrieval, Univ. of Minnesota, Minneapolis, Nov. 10-13, 1965, Ed. W. Simonton and C. Mason, pp. 87-97 (Univ. of Minnesota, Minneapolis, 1966).

Leech, P.C. and R.C. Matlack, Jr., Information Retrieval: Dictionary Representations and Cluster Evaluation, in Information Storage and Retrieval, Rept. No. ISR-12 [G. Salton, Proj. Dir.] pp. VII-1 to VII-20 (Cornell Univ., Ithaca, N.Y., June 1967).

Lees, R.B., Automatic Generation of Natural-Language Sentences, in Mathematical Machines and Their Applications, Colloq. Foundation of Mathematics, Tihany, 1962, pp. 177-184 (Ahed. Kiadó, Budapest, 1965).

Lefkovitz, D., The Application of the Digital Computer to the Problem of a Document Classification System, Proc. National Colloquium on Information Retrieval, Philadelphia, Pa., April 24-25, 1964, Ed. B.F. Cheydleur, pp. 133-146 (Spartan Books, Washington, D.C., 1965).

Lefkovitz, D., Substructure Search in the MCC System, J. Chem. Doc. 8, No. 3, 166-173 (Aug. 1968).

Lefkovitz, D. and N.S. Prywes, Automatic Stratification of Information, AFIPS Proc. Spring Joint Computer Conf., Vol. 23, Detroit, Mich., May 1963, pp. 229-240 (Spartan Books, Baltimore, Md., 1963).

Lehmann, W.P., Computational Linguistics: Procedures and Problems, Rept. No. LRC-65-WA-1, 1 v. (Linguistics Research Center, Texas Univ., Austin, Jan. 1965).

Lehmann, W.P. and E.D. Pendergraft, Development of a Linguistic Computer System, 85 p. (Linguistics Research Center, Univ. of Texas, Austin, June 1963).

Lehman, W.P. and E.D. Pendergraft, Development of a Linguistic Computer System, 33 p. (Linguistics Research Center, Univ. of Texas, Austin, Aug. 1964).

Lehmann, W.P. and E.D. Pendergraft, Machine Language Translation Study, Quarterly Progress Rept. No. 11, Nov. 1, 1965 - Jan. 31, 1966, 121 p. (Linguistics Research Center, Univ. of Texas, Austin, May 1966).

Lemmon, A., Automatic Identification of Phrases for Document Classification, in Information Storage and Retrieval, Rept. No. ISR-5 [G. Salton, Proj. Dir.] pp. II-1 to II-27 (Computation Lab., Harvard Univ., Cambridge, Mass., Jan. 1964).

LeSchack, A.R., The Determination of Clusters by Matrix Analysis, in Information Storage and Retrieval, Rept. No. ISR-7 [G. Salton, Proj. Dir.] pp. XIV-1 to XIV-43 (Computation Lab., Harvard Univ., Cambridge, Mass., June 1964).

LeSchack, A.R., A Note on Measures of Similarity, in Information Storage and Retrieval, Rept. No. ISR-7 [G. Salton, Proj. Dir.] pp. XIII-1 to XIII-16 (Computation Lab., Harvard Univ., Cambridge, Mass., June 1964).

Lesk, M.E., Operating Instructions for the SMART Text Processing and Document Retrieval System, in Information Storage and Retrieval, Rept. No. ISR-11 [G. Salton, Proj. Dir.] pp. II-1 to II-63 (Cornell Univ., Ithaca, N.Y., June 1966).

Lesk, M.E., Performance of Automatic Information Systems, Inf. Stor. & Retr. 4, No. 2, 201-218 (June 1968).

Lesk, M.E., Word-Word Associations in Document Retrieval Systems, in Information Storage and Retrieval, Rept. No. ISR-13 [G. Salton, Proj. Dir.] pp. IX-1 to IX-52 (Cornell Univ., Ithaca, N.Y., Jan. 1968).

Lesk, M.E., Word-Word Associations in Document Retrieval Systems, Am. Doc. 20, No. 1, 27-38 (Jan. 1969).

Lesk, M.E. and G. Salton, Design Criteria for Automatic Information Systems, in Information Storage and Retrieval, Rept. No. ISR-11 [G. Salton, Proj. Dir.] pp. V-1 to V-38 (Cornell Univ., Ithaca, N.Y., June 1966).

Lesk, M.E. and G. Salton, Interactive Search and Retrieval Methods Using Automatic Information Displays, in Information Storage and Retrieval, Rept. No. ISR-14 [G. Salton, Proj. Dir.] pp. IX-1 to IX-36 (Cornell Univ., Ithaca, N.Y., Oct. 1968).

Lesk, M.E. and G. Salton, Interactive Search and Retrieval Methods Using Automatic Information Displays, AFIPS Proc. Spring Joint Computer Conf., Vol. 34, Boston, Mass., May 14-16, 1969, pp. 435-446 (AFIPS Press, Montvale, N.J., 1969).

Lesk, M.E. and G. Salton, Relevance Assessments and Retrieval System Evaluation, in Information Storage and Retrieval, Rept. No. ISR-14 [G. Salton, Proj. Dir.] pp. III-1 to III-38 (Cornell Univ., Ithaca, N.Y., Oct. 1968).

Levery, F., An Automatic Indexing Experiment, [in French], Proc. 3rd AFCALTI Congress of Computing and Information Processing, Toulouse, France, May 1963, pp. 225-231 (Dunod, Paris, 1965).

Levien, R.E., Relational Data File: Experience With a System for Propositional Data Storage and Inference Execution, Research Memo. No. RM-5947-PR, 27 p. (The RAND Corp., Santa Monica, Calif., April 1969).

Levien, R. and M.E. Maron, Relational Data File: A Tool for Mechanized Inference Execution and Data Retrieval, Research Memo. No. RM-4793-PR, 89 p. (The RAND Corp., Santa Monica, Calif., Dec. 1965).

Lewis, P.A.W., P.B. Baxendale and J.L. Bennett, Statistical Discrimination of the Synonymy/Antonymy Relationship Between Words, J. ACM 14, No. 1, 20-44 (Jan. 1967).

Libbey, M.A., The Use of Second Order Descriptors for Document Retrieval, Am. Doc. 18, No. 1, 10-20 (Jan. 1967).

Licklider, J.C.R., Libraries of the Future, 219 p. (M.I.T. Press, Cambridge, Mass., 1965).

Licklider, J.C.R., Man-Computer Interaction in Information Systems, in Toward a National Information System, Second Annual National Colloquium on Information Retrieval, Philadelphia, Pa., April 23-24, 1965, Ed. M. Rubinoff, pp. 63-75 (Spartan Books, Washington, D.C., 1965).

Lieberman, D., Studies in Automatic Language Processing, 110 p. (IBM Corp., Thomas J. Watson Research Center, Yorktown Heights, N.Y., July 1965).

Lieberman, D., D. Lochak, N. Metas, K. Ochel and M. Carlson, Automatic Deep Structure Analysis Using an Approximate Formalism, Part I of [D. Lieberman], Studies in Automatic Language Processing, pp. 3-73 (IBM Corp., Thomas J. Watson Research Center, Yorktown Heights, N.Y., July 1965).

Lindzey, G. and E. Aronson, Eds., The Handbook of Social Psychology, 2nd ed., Vol. 2, Research Methods, 819 p. (Addison Wesley Pub. Co., Reading, Mass., 1968).

Lipetz, B.A., The Continuity Index of Documentation Abstracts, Proc. IFIP Congress 68, Edinburgh, Scotland, Aug. 5-10, 1968, Booklet G, pp. G 10 - G 12 (North-Holland Pub. Co., Amsterdam, 1968).

Lipetz, B.A., The Effect of a Citation Index on Literature Use by Physicists, Proc. 1965 Congress, F.I.D., 31st Meeting and Congress, Vol. II, Washington, D.C., Oct. 7-16, 1965, pp. 107-115 (Spartan Books, Washington, D.C., 1966).

Little, Arthur D., Inc., An Evaluation of Machine-Aided Translation Activity at FID, 74 p. (Cambridge, Mass., May 1, 1965).

Ljudskanov, A. and E. Paskaleva, A Possible Method of Reducing Stem Homonymy in the Automatic Analysis of Russian Text for Machine Translation, [in Russian], Computational Linguistics, No. 5, 118-121 (1966).

Locke, W.N. and A.D. Booth, Eds., Machine Translation of Languages, 243 p. (Pub. jointly by The Technology Press, M.I.T., and John Wiley, New York, 1955).

Lockheed Missiles & Space Company, Electronic Sciences Lab., Automatic Indexing and Abstracting, Annual Progress Report, Pt. 1, Rept. No. M-21-66-1, 1 vol. (Palo Alto, Calif., Mar. 1966).

Lockheed Missiles & Space Company, Electronic Sciences Lab., Automatic Indexing and Abstracting, Pt. 2: English Indexing of Russian Technical Text, Annual Progress Report, Rept. No. M-21-66-2, 72 p. (Palo Alto, Calif., Mar. 1966).

Lockheed Missiles & Space Company, Annual Report: Automatic Indexing and Abstracting, Rept. No. M-21-67-1, 118 p. (Sunnyvale, Calif., Mar. 1967).

Long, J.M., H.J. Barnhard and G.C. Levy, Dictionary Buildup and Stability of Word Frequency in a Specialized Medical Area, Am. Doc. 18, No. 1, 21-25 (Jan. 1967).

Lorr, M. and S.B. Lyerly, Conference on Cluster Analysis of Multivariate Data, New Orleans, La., Dec. 9-11, 1966, 323 p. (Catholic Univ. of America, Washington, D.C., June 1967).

Loukopoulos, L., Indexing Problems and Some of Their Solutions, Am. Doc. 17, No. 1, 17-25 (Jan. 1966).

Loveman, D.B., C.A. Moyne and R.G. Tobey, CUE: A Customized System for Restricted, Natural English, in IBM Proc. Inf. Systems Symp., Washington, D.C., Sept. 4-6, 1968, pp. 203-229 (IBM Corp., 1968).

Luce, R.D., R.R. Bush and E. Galanter, Eds., Readings in Mathematical Psychology, Vol. 2, 568 p. (Wiley, New York, 1965).

Lunin, L.F., M.T. Heath and R.H. Shepard, Differences Between Vocabularies Used in Pre- and Post-Research Documents. Preliminary Observations, in Levels of Interaction Between Man and Information, Proc. Am. Doc. Inst. Annual Meeting, Vol. 4, New York, N.Y., Oct. 22-27, 1967, pp. 137-141 (Thompson Book Co., Washington, D.C., 1967).

Lustig, G., The Development of an Automatic Indexing System at EURATOM, [Manuscript for 5th EURATOM-sponsored meeting of librarians working in the nuclear field, April 24-25, 1968; will be published as EURATOM report].

Lustig, G., A New Class of Association Factors, in Mechanized Information Storage, Retrieval and Dissemination, Proc. F.I.D./I.F.I.P. Joint Conf., Rome, Italy, June 14-17, 1967, Ed. K. Samuelson, pp. 213-224 (North-Holland Pub. Co., Amsterdam, 1968).

Lynch, M.F., Subject Indexes and Automatic Document Retrieval, J. Doc. 22, No. 3, 167-185 (Sept. 1966).

Magnino, J.J. Jr., IBM Technical Information Retrieval Center - Normal Text Techniques, in Toward a National Information System, Second Annual National Colloquium on Information Retrieval, Philadelphia, Pa., April 23-24, 1965, Ed. M. Rubinoff, pp. 199-215 (Spartan Books, Washington, D.C., 1965).

Magnino, J.J. Jr., IBM's Unique but Operational International Industrial Textual Documentation System --- ITIRC, Paper III c 5, 33rd Conf. F.I.D. and Int. Cong. on Documentation, Tokyo, Sept. 12-22, 1967, 9 p. (preprint).

Maher, J.J., Ed., Proc. Workshop on Working with Semi-Automatic Documentation Systems, Warrenton, Va., May 2-5, 1965, 106 p. (System Development Corp., Santa Monica, Calif., 1965).

Maloney, C.J., Practical Preparation of Material Indexes, The Indexer 5, No. 2, 81-90 (Autumn 1966).

Maloney, C.J. and M.N. Epstein, Progress in Internal Indexing, in Progress in Information Science and Technology, Proc. Am. Doc. Inst. Annual Meeting, Vol. 3, Santa Monica, Calif., Oct. 3-7, 1966, pp. 57-62 (Adrianne Press, 1966).

Maloney, C.J., S. Bryan and M. Epstein, Computer Assisted Primary Index Preparation, J. Chem. Doc. 7, No. 4, 223-232 (Nov. 1967).

Maloney, C.J., J. Dukes and S. Green, Indexing Reports by Computer, in Colloquium on Technical Preconditions for Retrieval Center Operations, Philadelphia, Pa., April 1964, Ed. B.F. Cheydleur, pp. 13-28 (Spartan Books, Washington, D.C., 1965).

Mandersloot, W.G.B. and R. McGillivray, Documentation and Information Retrieval: Application of Keyword Indexing, WNNR CSIR Spec. Rep. CHEM 38, 1965.

Markuson, B.E., Automation in Libraries and Information Centers, in Annual Review of Information Science and Technology, Vol. 2, Ed. C.A. Cuadra, pp. 255-284 (Interscience Pub., New York, 1967).

Maron, M.E., A Logician's View of Language Data Processing, in Natural Language and the Computer, Ed. P.L. Garvin, pp. 128-150 (McGraw-Hill, New York, 1963).

Maron, M.E., Mechanized Documentation: The Logic Behind a Probabilistic Interpretation, in Statistical Association Methods For Mechanized Documentation, Symp. Proc., Washington, D.C., Mar. 17-19, 1964, NBS Misc. Pub. 269, Ed. M.E. Stevens et al, pp. 9-13 (U.S. Govt. Print. Off., Washington, D.C., Dec. 15, 1965).

Martins, G.R. and S.B. Smith, BEAST: The Basic Experimental Automatic Syntactic Translator, 117 p. (Bunker-Ramo Corp., Canoga Park, Calif., June 1, 1967).

Martins, G.R. and S.B. Smith, Computer-Aided Research in Machine Translation D199. A Parsing Procedure for a Vector-Symbol Phrase Grammar of Russian, 97 p. (Bunker-Ramo Corp., Canoga Park, Calif., Dec. 1965).

Martyn, J., Citation Indexing, Indexer 5, 5-15 (Spring 1966).

Martyn, J., An Examination of Citation Indexes, Aslib Proc. 17, 184-196 (1965).

Masterman, M., Man-Aided Computer Translation from English into French Using an On-Line System to Manipulate a Bilingual Conceptual Dictionary or Thesaurus, in Second Conférence Internationale sur le Traitement Automatique des Langues, Grenoble, Aug. 1967, Paper No. 12.

Mathews, W. D., The TIP Retrieval System at MIT, in Information Retrieval - A Critical View, Based on Third Annual Colloquium on Information Retrieval, Philadelphia, Pa., May 12-13, 1966, Ed. G. Schecter, pp. 95-108 (Thompson Book Co., Washington, D.C., 1967).

Matthews, G.M. and J. VanLuik, Citation and Subject Indexing in Science, Lib. Res. & Tech. Serv. 9, 478-482 (1965).

McConolgue, K. and R.F. Simmons, Analyzing English Syntax with a Pattern-Learning Parser, Commun. ACM 8, 687-698 (Nov. 1965).

McDonald, R.R., Linguistic Structure, in Information Systems Compatibility, Ed. S.M. Newman, pp. 85-102 (Spartan Books, Washington, D.C., 1965).

McNamara, J. and D. Stone, Machine Extracting Progress, in Progress in Information Science and Technology, Proc. Am. Doc. Inst. Annual Meeting, Vol. 3, Santa Monica, Calif., Oct. 3-7, 1966, pp. 217-223 (Adrianne Press, 1966).

Meadow, C.T. and D.W. Waugh, Computer-Assisted Interrogation, AFIPS Proc. Fall Joint Computer Conf., Vol. 29, San Francisco, Calif., Nov. 7-10, 1966, pp. 381-394 (Spartan Books, Washington, D.C., 1966).

Meetham, A.R., Graph Separability and Word Grouping, Proc. 21st National Conf., ACM, Los Angeles, Calif., Aug. 30 - Sept. 1, 1966, pp. 513-514 (Thompson Book Co., Washington, D.C., 1966).

Meetham, A.R., Probabilistic Pairs and Groups of Words in a Text, Language & Speech 7, 98-106 (1964).

Melton, J.S., Automatic Language Processing for Information Retrieval: Some Questions, in Progress in Information Science and Technology, Proc. Am. Doc. Inst. Annual Meeting, Vol. 3, Santa Monica, Calif., Oct. 3-7, 1966, pp. 255-263 (Adrianne Press, 1966).

Melton, J.S., Automatic Processing of Metallurgical Abstracts for the Purpose of Information Retrieval, Interim Rept. NSF-1, Jan. 1963, 65 p.; Interim Rept. NSF-2, Feb. 1964, 104 p.; Interim Rept. NSF-3, July 1965, 2 Vols., 413 p.; Final Rept. NSF-4, July 1, 1967, 95 p. (Case Western Reserve Univ., Cleveland, O., 1963-1967).

Melton, J.S., Major Contemporary Topics in Documentation, in Information in the Language Sciences, Proc. Conf. on Information in the Language Sciences, Warrenton, Va., Mar. 4-6, 1966, sponsored by the Center for Applied Linguistics, Ed. R.R. Freeman et al, pp. 69-93 (American Elsevier Pub. Co., New York, 1968).

Melton, J.S., A Use for the Techniques of Structural Linguistics in Documentation Research, Rept. No. CSL:TR-4, 20 p. (Center for Documentation and Communication Research, Western Reserve Univ., Cleveland, O., Sept. 1964). Also in Proc. Second Int. Study Conf. on Classification Research, Elsinore, Denmark, Sept. 14-18, 1964, Ed. P. Atherton, pp. 466-480 (Munksgaard, Copenhagen, 1965).

Mikhailov, A.I., Studies on Automatic Indexing and Abstracting in the USSR, Working paper for the Unesco-VINITI Symp. on Mechanized Abstracting and Indexing, Moscow, Sept. 28 - Oct. 1, 1966, 12 p. Also in Symp. on Mechanized Abstracting and Indexing - Papers and Discussion, Invited Paper, pp. 33-41, Unesco Doc. No. SC/WS/172, issued Paris, Jan. 12, 1968 (Distribution Limited).

Miller, L., J. Minker, W.G. Reed and W.E. Shindle, A Multi-Level File Structure for Information Processing, Proc. Western Joint Computer Conf., Vol. 17, San Francisco, Calif., May 3-5, 1960, pp. 53-59 (Pub. by WJCC, San Francisco, Calif., 1960).

Minsky, M.L., Artificial Intelligence, Scient. Amer. 215, No. 3, 247-260 (Sept. 1966).

Minsky, M., Matter, Mind and Models, in Information Processing 1965, Proc. IFIP Congress 65, Vol. 1, New York, N.Y., May 24-29, 1965, Ed. W.A. Kalenich, pp. 45-49 (Spartan Books, Washington, D.C., 1965).

Minsky, M., A Selected Descriptor-Indexed Bibliography to the Literature on Artificial Intelligence, in Computers and Thought, Ed. E.A. Feigenbaum and J. Feldman, pp. 453-523 (McGraw-Hill, New York, 1963).

Minsky, M., Steps Toward Artificial Intelligence, in Computers and Thought, Ed. E.A. Feigenbaum and J. Feldman, pp. 406-452 (McGraw-Hill, New York, 1963).

Montague, B.A., Testing Comparison, and Evaluation of Recall, Relevance, and Cost of Coordinate Indexing with Links and Roles, Am. Doc. 16, No. 3, 201-208 (July 1965).

Mooers, C.N., The Indexing Language of an Information Retrieval System, in Information Retrieval Today, papers presented at the Institute conducted by the Library School and the Center for Continuation Study, Univ. of Minnesota, Minneapolis, Sept. 19-22, 1962, Ed. W. Simonton, pp. 21-36 (Univ. of Minnesota, Minneapolis, 1963).

Moravcova, V., Automatizace Referování v Odbourné Literature (Automatic Abstracting of Literature), Metodika a Technika Informaci, Nos. 6-7, 17-20 (1967).

Morenoff, E. and J.B. McLean, Classifier: An Automated Computer-Oriented Information Classification System, in Parameters of Information Science, Proc. Am. Doc. Inst. Annual Meeting, Vol. 1, Philadelphia, Pa., Oct. 5-8, 1964, pp. 411-420 (Spartan Books, Washington, D.C., 1964).

Motherwell, G.M., CONDEX: A Technique for Automatic Sub-Documentary Indexing, in Progress in Information Science and Technology, Proc. Am. Doc. Inst. Annual Meeting, Vol. 3, Santa Monica, Calif., Oct. 3-7, 1966, pp. 299-306 (Adrianne Press, 1966).

Moureau, M. and J.M. Lasvergeres, Automatic Indexing of IFP Scientific and Technical Reports, in Mechanized Information Storage, Retrieval and Dissemination, Proc. F.I.D./I.F.I.P. Joint Conf., Rome, Italy, June 14-17, 1967, Ed. K. Samuelson, pp. 468-484 (North-Holland Pub. Co., Amsterdam, 1968).

Murdock, J.W. and D.M. Liston, Jr., A General Model of Information Transfer: Theme Paper 1968 Annual Convention, Am. Doc. 18, No. 4, 197-208 (Oct. 1967).

National Academy of Sciences-National Research Council, Language and Machines: Computers in Translation and Linguistics, Pub. 1416, 124 p. (NAS-NRC, Washington, D.C., 1966).

Needham, R.M., Applications of the Theory of Clumps, Mech. Trans. 8, 113-127 (1965).

Needham, R.M., Information Retrieval and Some Cognate Computing Problems, in Advances in Programming and Non-Numerical Computation, Ed. L. Fox, pp. 201-218 (Pergamon, New York, 1966).

Needham, R.M., Problems of Scale in Automatic Classification, (Abstract), in Statistical Association Methods For Mechanized Documentation, Symp. Proc., Washington, D.C., Mar. 17-19, 1964, NBS Misc. Pub. 269, Ed. M.E. Stevens et al, p. 157 (U.S. Govt. Print. Off., Washington, D.C., Dec. 15, 1965).

Needham, R.M., Semantic Problems of Machine Translation, in Information Processing 1965, Proc. IFIP Congress 65, Vol. 1, New York, N.Y., May 24-29, 1965, Ed. W.A. Kalenich, pp. 65-69 (Spartan Books, Washington, D.C., 1965).

Needham, R.M. and K. Sparck-Jones, Keywords and Clumps. Recent Work on Information Retrieval at the Cambridge Language Research Unit, J. Doc. 20, No. 1, 5-15 (Mar. 1964).

Neelameghan, A., The Keyword-in-Context Index: An Evaluation, in Documentation Periodicals: Coverage, Arrangement, Scatter, Seepage, Compilation, Ed. S.R. Ranganathan and A. Neelameghan, pp. 127-140 (Documentation Research and Training Center, Bangalore, India, 1963).

Nelson, P.J., User Profiling for Normal Text Retrieval, in Levels of Interaction Between Man and Information, Proc. Am. Doc. Inst. Annual Meeting, Vol. 4, New York, N.Y., Oct. 22-27, 1967, pp. 288-295 (Thompson Book Co., Washington, D.C., 1967).

Newcomb, M.A. and R.A. Benson, Technique for the Automatic Generation of Bibliographies, (Biomedical Information Application), 2 Vols. (General Dynamics/Convair, San Diego, Calif., 1965).

Newell, V.A. and W. Goffman, Searching Titles by Man, Machine, and Chance, in Parameters of Information Science, Proc. Am. Doc. Inst. Annual Meeting, Vol. 1, Philadelphia, Pa., Oct. 5-8, 1964, pp. 421-423 (Spartan Books, Washington, D.C., 1964).

Newman, S.M., Ed., Information Systems Compatibility, based on papers given at the 6th Institute on Information Storage and Retrieval, American Univ., Washington, D.C., 1964, 150 p. (Spartan Books, Washington, D.C., 1965).

O'Connor, J., Automatic Subject Recognition in Scientific Papers: An Empirical Study, J. ACM 12, No. 4, 490-515 (Oct. 1965).

O'Connor, J., Letter to the Editor, Am. Doc. 19, No. 1, 104 (Jan. 1968).

O'Connor, J., Mechanized Indexing Methods and Their Testing, J. ACM 11, No. 4, 437-449 (Oct. 1964).

O'Connor, J., Mechanized Indexing Studies of MSD Toxicity, Part I and Part II, Rept. No. AFOSR-64-0682, 128 p. (Univ. of Pennsylvania, Philadelphia and the Institute for Scientific Information, Philadelphia, Pa., Jan. 1964).

O'Connor, J., Methods of Mechanized Indexing: Comprehensive Document Preparation and Testing Mechanized Indexing Quality, 29 p. (Inst. for Cooperation Research, Univ. of Pennsylvania, Philadelphia, 1963).

Oettinger, A.G., Automatic Processing of Natural and Formal Languages, in Information Processing 1965, Proc. IFIP Congress 65, Vol. 1, New York, N.Y., May 24-29, 1965, Ed. W.A. Kalenich, pp. 9-16 (Spartan Books, Washington, D.C., 1965).

Oettinger, A.G., Computational Linguistics, Am. Math. Monthly 72, Part II, 147-150 (Feb. 1965).

Oettinger, A.G., Ed., Mathematical Linguistics and Automatic Translation, Rept. No. NSF-15 to the National Science Foundation, 1 v. (Computation Lab., Harvard Univ., Cambridge, Mass., Aug. 1965).

Oettinger, A.G., Mathematical Linguistics and Automatic Translation, 285 p. (Computation Lab., Harvard Univ., Cambridge, Mass., Dec. 1965).

Oettinger, A.G., Mathematical Linguistics and Automatic Translation, 294 p. (Harvard Univ., Cambridge, Mass., Aug. 1966).

Ofer, K.D., A Computer Program to Index or Search Linear Notations, J. Chem. Doc. 8, No. 3, 128-129 (Aug. 1968).

Ohringer, L., Accumulation of Natural Language Text for Computer Manipulation, in Parameters of Information Science, Proc. Am. Doc. Inst. Annual Meeting, Vol. 1, Philadelphia, Pa., Oct. 5-8, 1964, pp. 311-313 (Spartan Books, Washington, D.C., 1964).

Orr, D.B. and V.H. Small, Comprehensibility of Machine-Aided Translations of Russian Scientific Documents, Mechanical Translation and Computational Linguistics 10, Nos. 1 & 2, 1-10 (Mar. and June 1967).

Ossorio, P.G., Dissemination Research, Rept. No. RADC-TR-65-314, 75 p. (Rome Air Development Center, Griffiss AFB, New York, Dec. 1965).

Overmyer, L., The Evolution of a Computer-Produced Index for Diabetes Literature, in Progress in Information Science and Technology, Proc. Am. Doc. Inst. Annual Meeting, Vol. 3, Santa Monica, Calif., Oct. 3-7, 1966, pp. 307-320 (Adrianne Press, 1966).

Overton, R.K., Intelligent Machines and Hazy Questions, Computers & Automation 14, No. 7, 26-30 (1965).

PANDEX. Current Index to Scientific and Technical Literature, Index series issued periodically in printed, microfiche, and magnetic tape form, CCM Information Sciences Inc., New York (1967 et seq.).

Parsons, C.D., SIR: A Statistical Information Retrieval System, Proc. 19th National Conf., ACM, Philadelphia, Pa., Aug. 25-27, 1964, pp. L2.1-1 to L2.1-7 (Assoc. for Computing Machinery, New York, 1964).

Pendergraft, E.D., Translating Languages, in Automated Language Processing, Ed. H. Borko, pp. 291-323 (Wiley, New York, 1967).

Pendergraft, E. and N. Dale, Automatic Linguistic Classification, 46 p. (Linguistic Research Center, Univ. of Texas, Austin, Nov. 1965).

Perschke, S., Automatic Language Translation, Its Possibilities and Limitations, Euratom Bulletin, No. 2, 1967, 8 p.

Perschke, S., Machine Translation --- The Second Phase of Development, Endeavour 27, No. 101, 97-100 (May 1968).

Perschke, S., The Use of Machine Translation in Documentation [Manuscript for 5th EURATOM-sponsored meeting of librarians working in the nuclear field, April 24-25, 1968; will be published as EURATOM report], 13 p.

Perschke, S., The Use of the "SLC" System in Automatic Indexing, in Mechanized Information Storage, Retrieval and Dissemination, Proc. F.I.D./I.F.I.P. Joint Conf., Rome, Italy, June 14-17, 1967, Ed. K. Samuelson, pp. 300-311 (North-Holland Pub. Co., Amsterdam, 1968).

Peters, S., Evaluation of Retrieval Systems, Appendix III of Centralization and Documentation, Final Report to the National Science Foundation, pp. 71-109 (Arthur D. Little, Inc., Cambridge, Mass., July 1963).

Petrarca, A.E. and W.M. Lay, The Double-KWIC Coordinate Index. A New Approach for Preparation of High-Quality Printed Indexes by Automatic Indexing Techniques, presented in part before the Division of Chemical Literature, 157th Meeting, ACS, Minneapolis, Minn., April 1969, 29 p. (Dept. of Computer and Information Science, The Ohio State Univ., Columbus, 1969).

Petrarca, A.E. and W.M. Lay, The Double KWIC Coordinate Index. II. Use of an Automatically Generated Authority List to Eliminate Scattering Caused by Some Singular and Plural Main Index Terms, Tech. Rept. 69-9, 13 p. (Computer and Information Science Research Center, The Ohio State Univ., Columbus, Aug. 1969).

Petrick, S.R., A LISP Program for the Parsing of Sentences with Respect to a Transformational Grammar, in Information Processing 1965, Proc. IFIP Congress 65, Vol. 2, Ed. W.A. Kalenich, pp. 528-529 (Spartan Books, Washington, D.C., 1966).

Phillips, A.V., A Question-Answering Routine, Master's Thesis, RLE and MIT Computation Center Memo 16, 20 p. (M.I.T., Cambridge, Mass., May 21, 1960).

Picot, G., M.L. Deribere-Desgardes and F. Levery, Experiment in the Automatic Selection of Documentation, [Partial trans. (p. 8-11) of Revue de la Documentation 29, No. 1, 8-13 (1962)], 1962, 10 p.

Post, P.B., A Lifelike Model for Associative Relevance, Proc. Int. Joint Conf. on Artificial Intelligence, Washington, D.C., May 7-9, 1969, Ed. D.E. Walker and L.M. Norton, pp. 271-280 (Int. Joint Conf. on Artificial Intelligence, 1969).

Potocko, R.J., An Indexed System of Terms with Related and Associated Words, Am. Doc. 19, No. 2, 146-150 (April 1968).

Preparata, F.P. and R.T. Chien, On Clustering Techniques of Citation Graphs, Rept. No. R-349 (Coordinated Science Lab., Univ. of Illinois, Urbana, May 1967).

Press, L.I. and M.S. Rogers, IDEA - A Conversational, Heuristic Program for Inductive Data Exploration and Analysis, Proc. 22nd National Conf., ACM, Washington, D.C., Aug. 29-31, 1967, pp. 35-40 (Thompson Book Co., Washington, D.C., 1967).

Price, N. and S. Schiminovich, A Clustering Experiment: First Step Towards a Computer-Generated Classification Scheme, Inf. Stor. & Retr. 4, No. 3, 271-280 (Aug. 1968).

Prywes, N.S., Man-Computer Problem Solving with Multilist, Proc. IEEE 54, 1788-1801 (Dec. 1966).

Prywes, N.S. and M. Silver, An Information Center for Effective R & D Management, in Second Cong. on the Information System Sciences, Hot Springs, Va., Nov. 1964, Ed. J. Spiegel and D.E. Walker, pp. 105-116 (Spartan Books, Washington, D.C., 1965).

Psathas, G., The General Inquirer: Useful or Not?, Computers and the Humanities 3, No. 3, 163-174 (Jan. 1969).

Pylyshyn, Z.W., FINDSIT: A Computer Program for Language Research, Behav. Sci. 14, No. 3, 248-251 (May 1969).

Quillian, M.R., The Teachable Language Comprehender: A Simulation Program and Theory of Language, Commun. ACM 12, No. 8, 459-476 (Aug. 1969).

Raizada, A.S., L.J. Haravu and S.N. Sur, Directory Compilation by Computer, Annals of Lib. Sci. 14, No. 2, 89-101 (June 1967).

Ranganathan, S.R. and A. Neelameghan, Eds., Documentation Periodicals: Coverage, Arrangement, Scatter, Seepage, Compilation, 244 p. (Documentation Research and Training Center, Bangalore, India, 1963).

Raphael, B., Aspects and Applications of Symbol Manipulation, Proc. 21st National Conf., ACM, Los Angeles, Calif., Aug. 30 - Sept. 1, 1966, pp. 69-74 (Thompson Book Co., Washington, D.C., 1966).

Raphael, B., SIR: A Computer Program for Semantic Information Retrieval, Rept. No. MAC-TR-2, 169 p. (M.I.T., Cambridge, Mass., June 1964).

Reed, D.M. and D.J. Hillman, Document Retrieval Theory, Relevance, and the Methodology of Evaluation, Rept. No. 3, Microcategorization for Text Processing, 41 p. (Center for the Information Sciences, Lehigh Univ., Bethlehem, Pa., July 7, 1966).

Reed, D.M. and D.J. Hillman, Document Retrieval Theory, Relevance, and the Methodology of Evaluation, Rept. No. 4, Canonical Decomposition, 33 p. (Center for the Information Sciences, Lehigh Univ., Bethlehem, Pa., Aug. 1966).

Regnier, S., Sur Quelques Aspects Mathématiques des Problèmes de Classification Automatique, (On Some Problems of Automatic Classification, ICC Bull. 4, No. 3, 175-191 (July-Sept. 1965).

Reichling, A., Possibilities and Limitations of Mechanical Translation, From the Linguist's Point of View, Sprachkunde und Informationsverarbeitung, No. 1, 23-32 (1963), [German].

Reifler, E., The Mechanical Determination of Meaning, in Machine Translation of Languages, Ed. W.N. Locke and A.D. Booth, pp. 136-164 (Pub. jointly by The Technology Press, M.I.T., and John Wiley, New York, 1955).

Reintjes, J.F. and R.S. Marcus, Computer-Aided Processing of the News, AFIPS Proc. Spring Joint Computer Conf., Vol. 34, Boston, Mass., May 14-16, 1969, pp. 331-338 (AFIPS Press, Montvale, N.J., 1969).

Reisner, P., Semantic Diversity and a "Growing" Man-Machine Thesaurus, in Some Problems in Information Science, Ed. M. Kochen, pp. 117-130 (The Scarecrow Press, Inc., New York, 1965).

Reitsma, K. and J. Sagalyn, Correlation Measures, in Information Storage and Retrieval, Rept. No. ISR-13 [G. Salton, Proj. Dir.] pp. IV-1 to IV-30 (Cornell Univ., Ithaca, N.Y., Jan. 1968).

Reitz, G. and W.L. White, A Console System for Text Manipulation and Page Composition. Computer-Aided Research in Machine Translation, Progress Report No. 13, 22 p. (Bunker-Ramo Corp., Canoga Park, Calif., Mar. 20, 1967).

Resnick, A. and T.R. Savage, The Consistency of Human Judgements of Relevance, in The Coming Age of Information Technology, pp. 57-62 (Documentation, Inc., Bethesda, Md., 1965).

Revard, C., On the Computability of Certain Monsters in Noah's Ark: Using Computers to Study Webster's Seventh New Collegiate Dictionary and The New Merriam-Webster Pocket Dictionary, Proc. 23rd National Conf., ACM, Las Vegas, Nev., Aug. 27-29, 1968, pp. 807-813 (Brandon/Systems Press, Inc., Princeton, N.J., 1968).

Revzin, I.I., Puti Preodoleniya Krizisa v Vychislitel'noi Lingvistike (2-ya Mezhdunarodnaya Konferentsiya po Automatizatsii Linguisticheskikh), (Toward Overcoming the Crisis in Computational Linguistics), 2nd International Conf. for Automation of Linguistic Research, Grenoble, France, Aug. 23-25, 1967, Nauchno-Tekhnicheskaya Informatsiya, Series 2 (2) 39-42 (1968). [in Russian].

Rhodes, I., The Method for Mechanical Translation Used by the National Bureau of Standards Group and the Structure of Its Machine Glossary, in Automation and Scientific Communication, Short Papers, Pt. 1, papers contributed to the Theme Sessions of the 26th Annual Meeting, Am. Doc. Inst., Chicago, Ill., Oct. 6-11, 1963, Ed. H.P. Luhn, pp. 23-24 (Am. Doc. Inst., Washington, D.C., 1963).

Richmond, P.A., Transformation and Organization of Information Content: Aspects of Recent Research in the Art and Science of Classification, Proc. 1965 Congress F.I.D., 31st Meeting and Congress, Vol. II, Washington, D.C., Oct. 7-16, 1965, pp. 87-106 (Spartan Books, Washington, D.C., 1966).

Rioch, D.McK. and E.A. Weinstein, Eds., Disorders of Communication: Res. Publ. Assoc. Res. Nerv. Ment. Dis. Vol. 42 (Williams and Wilkins, Baltimore, Md., 1964).

Ripperger, E.A., H. Wooster and S. Juhasz, WADEX (Word Author inDEX) A New Tool in Literature Retrieving, Mech. Engr. 86, No. 3, 45-50 (Mar. 1964).

Ripperger, E.A., H. Wooster, S. Juhasz and D. Falconer, Applied Mechanics Reviews, WADEX Word and Author Index, Volume XVI, 1963, Rept. No. AFOSR-65-0728, 627 p. (Southwest Research Inst., San Antonio, Texas, Sept. 1965).

Ripperger, E.A., H. Wooster, S. Juhasz and D. Falconer, Second Experimental WADEX for AMR Vol. 16, 650 p. (U.S. Govt. Print. Off., Washington, D.C., July 1965).

Ripperger, E.A., H. Wooster, S. Juhasz, D. Falconer, F.A. Roach and P.L. Stanton, WADEX, Word and Author Index, Part I - Description; Part II - Program Documentation, AMR Rept. No. 38, 165 p. (Applied Mechanics Reviews, San Antonio, Texas, Mar. 1966).

Rittenhouse, J.O. Jr., Indexing by Computer (LITE), AF JAG L. Rev. 8, No. 6, 26-35 (Nov.-Dec. 1966).

Robinson, H.R., Annual Report: Automatic Indexing and Abstracting, Part II: English Indexing of Russian Technical Texts, 1 v. (Lockheed Missiles & Space Co., Palo Alto, Calif., 1966).

Robinson, J.J., The Transformation of Sentences for Information Retrieval, Proc. 1965 Congress F.I.D., 31st Meeting and Congress, Vol. II, Washington, D.C., Oct. 7-16, 1965, pp. 69-73 (Spartan Books, Washington, D.C., 1966).

Robinson, J. and S. Marks, PARSE: A System for Automatic Syntactic Analysis of English Text. Part 1, Rept. No. RM-4654-PR, Pt. 1, 204 p. (The RAND Corp., Santa Monica, Calif., Sept. 1965).

Robinson, J. and S. Marks, PARSE: A System for Automatic Syntactic Analysis of English Text. Part 2, Rept. No. RM-4654-PR, Pt. 2, 267 p. (The RAND Corp., Santa Monica, Calif., Sept. 1965).

Rocchio, J.J. Jr., Document Retrieval Systems --- Optimization and Evaluation, in Information Storage and Retrieval, Rept. No. ISR-10 [G. Salton, Proj. Dir.] 1 vol. (Harvard Univ., Cambridge, Mass., Mar. 1966).

Rohrer, C., Automatic Analysis of a French Text, Linguistik und Informationsverarbeitung, No. 10, 50-65 (Feb. 1967).

Rolling, L.N., A Computer-Aided Information Service for Nuclear Science and Technology, EURATOM, CID, Brussels, Belgium, 1966, 29 p.

Rolling, L., Keyword Index for Machine Documentation in the Nuclear Field, ERU/CID/4243/10/62/e, EURATOM, Brussels, Belgium, Aug. 1962, 1 v.

Rolling, L., Un Repertoire de Mots-Cles pour la Documentation Mecanisee dans le Domaine de la Technique Nucleaire, Bull. Bibliotheques de France 8, 11-25 (1963).

Rolling, L. and J. Piette, Interaction of Economics and Automation in a Large-Size Retrieval System, in Mechanized Information Storage, Retrieval and Dissemination, Proc. F.I.D./I.F.I.P. Joint Conf., Rome, Italy, June 14-17, 1967, Ed. K. Samuelson, pp. 367-390 (North-Holland Pub. Co., Amsterdam, 1968).

Roper, D.C. and W.D. Timberlake, Computer-Produced Indexes, 18 p. (IBM Corp., Poughkeepsie, N.Y., 1963).

Rose, M.J., Classification of a Set of Elements, The Computer J. 7, 208-211 (Oct. 1964).

Rosenberg, K.C. and C.L.M. Blocher, A Comparison of the Relevance of Key-Word-in-Context versus Descriptor Indexing Terms, Am. Doc. 19, No. 1, 27-29 (Jan. 1968).

Rothman, J., Communicating with Indexes, Spec. Lib. 57, No. 8, 569-570 (Oct. 1966).

Roudabush, G.E., C.R.T. Bacon, R.P. Briggs, J.A. Fierst, D.W. Isner and H.A. Nogun, The Left Hand of Scholarship: Computer Experiments with Recorded Text as a Communication Media, AFIPS Proc. Fall Joint Computer Conf., Vol. 27, Pt. 1, Las Vegas, Nev., Nov. 30 - Dec. 1, 1965, pp. 399-411 (Spartan Books, Washington, D.C., 1965).

Rubenstein, H. and J.B. Goodenough, Contextual Correlates of Synonymy, Commun. ACM 8, No. 10, 627-633 (Oct. 1965).

Rubinoff, M., Ed., Toward a National Information System, Second Annual National Colloquium on Information Retrieval, Philadelphia, Pa., April 23-24, 1965, 242 p. (Spartan Books, Washington, D.C., 1965).

Rubinoff, M. and D.C. Stone, Semantic Tools in Information Retrieval, in Levels of Interaction Between Man and Information, Proc. Am. Doc. Inst. Annual Meeting, Vol. 4, New York, N.Y., Oct. 22-27, 1967, pp. 169-174 (Thompson Book Co., Washington, D.C., 1967).

Rudin, B.D., Some Approaches to Automatic Indexing, in Computer and Information Sciences-II, Proc. 2nd Symp. on Computer and Information Sciences, Columbus, O., Aug. 22-24, 1966, Ed. J.T. Tou, pp. 291-301 (Academic Press, New York, 1967).

Rusconi, F. and P. Terzi, The Study of the Enrichment of Phrases, A Means of Obtaining the Elements of Summaries and Abstracts, in Mechanized Information Storage, Retrieval and Dissemination, Proc. F.I.D./I.F.I.P. Joint Conf., Rome, Italy, June 14-17, 1967, Ed. K. Samuelson, pp. 247-255 (North-Holland Pub. Co., Amsterdam, 1968).

Rush, J.E., R. Salvador and A. Lamora, Automatic Abstracting, presented in part before the Division of Chemical Literature, 157th National Meeting, Am. Chem. Soc., Minneapolis, Minn., April 13-18, 1969, 67 p. (The Ohio State Univ., Columbus, June 12, 1969).

Russell, M. and R.R. Freeman, Computer-Aided Indexing of a Scientific Abstract Journal by the UDC with UNIDEK: A Case Study, UDC Project Rept. No. AIP/UDC-4, 20 p. (American Institute of Physics, New York, 1967).

Sackman, B.S., Electrifying Law Research, Law & Computer Tech. 1, No. 3, 2-6 (Mar. 1968).

Sagalovich, N.M., Machine Indexing of Abstract Journals, Transl. from Nauchno-Tekhnicheskaya Informatsiya, No. 6, 18-20 (1963). Foreign Develop. Mach. Transl. Inf. Proc. No. 149 JPRS: 22547 (Dec. 1963) Off. Tech. Serv., Washington, D.C., 32-42.

Sage, C.R., R.R. Anderson and D.R. Fitzwater, Adaptive Information Dissemination, Am. Doc. 16, No. 3, 185-200 (July 1965).

Salton, G., Automated Language Processing, in Annual Review of Information Science and Technology, Vol. 3, Ed. C.A. Cuadra, pp. 169-199 (Encyclopedia Britannica, Inc., Chicago, Ill., 1968).

Salton, G., Automatic Document Selection, [Letters to the Editor], J. Doc. 24, No. 2, 119 (June 1968).

Salton, G., Automatic Phrase Matching, in Readings in Automatic Language Processing, Ed. D.G. Hays, pp. 169-188 (American Elsevier Pub., New York, 1966).

Salton, G., A Comparison Between Manual and Automatic Indexing Methods, Rept. No. 68-11, 42 p. (Dept. of Computer Science, Cornell Univ., Ithaca, N.Y., Mar. 1968).

Salton, G., A Comparison Between Manual and Automatic Indexing Methods, Am. Doc. 20, No. 1, 61-71 (Jan. 1969).

Salton, G., An Evaluation Program for Associative Indexing, in Statistical Association Methods For Mechanized Documentation, Symp. Proc., Washington, D.C., Mar. 17-19, 1964, NBS Misc. Pub. 269, Ed. M.E. Stevens et al, pp. 201-210 (U.S. Govt. Print. Off., Washington, D.C., Dec. 15, 1965).

Salton, G., The Identification of Document Content: A Problem in Automatic Information Retrieval, Proc. Harvard Symp. on Digital Computers and Their Applications, Cambridge, Mass., April 1961, Annals of Computation Laboratory, Vol. 21, pp. 273-304 (Harvard Univ., Cambridge, Mass., 1962).

Salton, G., Information Dissemination and Automatic Information Systems, Proc. IEEE 54, 1663-1678 (Dec. 1966).

Salton, G., Proj. Dir., Information Storage and Retrieval, Scient. Rept. No. ISR-5 to the National Science Foundation, 1 v. (Computation Lab., Harvard Univ., Cambridge, Mass., Jan. 1964).

Salton, G., Proj. Dir., Information Storage and Retrieval, Scient. Rept. No. ISR-7 to the National Science Foundation, 1 v. (Computation Lab., Harvard Univ., Cambridge, Mass., June 1964).

Salton, G., Proj. Dir., Information Storage and Retrieval, Scient. Rept. No. ISR-10 to the National Science Foundation, 1 v. (Computation Lab., Harvard Univ., Cambridge, Mass., Mar. 1966).

Salton, G., Proj. Dir., Information Storage and Retrieval, Scient. Rept. No. ISR-11 to the National Science Foundation, 1 v. (Dept. of Computer Science, Cornell Univ., Ithaca, N.Y., June 1966).

Salton, G., Proj. Dir., Information Storage and Retrieval, Scient. Rept. No. ISR-12 to the National Science Foundation, Reports on Evaluation, Clustering, and Feedback, 1 v. (Dept. of Computer Science, Cornell Univ., Ithaca, N.Y., June 1967).

Salton, G., Proj. Dir., Information Storage and Retrieval, Scient. Rept. No. ISR-13 to the National Science Foundation, Reports on Evaluation Procedure and Results 1965-1967, 1 v. (Dept. of Computer Science, Cornell Univ., Ithaca, N.Y., Jan. 1968).

Salton, G., Proj. Dir., Information Storage and Retrieval, Scient. Rept. No. ISR-14 to the National Science Foundation, Reports on Analysis, Search and Iterative Retrieval, 1 v. (Dept. of Computer Science, Cornell Univ., Ithaca, N.Y., Oct. 1968).

Salton, G., Progress in Automatic Information Retrieval, IEEE Spectrum 2, No. 8, 90-103 (Aug. 1965).

Salton, G., Search and Retrieval Experiments in Real-Time Information Retrieval, in Information Storage and Retrieval, Rept. No. ISR-14, [G. Salton, Proj. Dir.] pp. VII-1 to VII-30 (Dept. of Computer Science, Cornell Univ., Ithaca, N.Y., Oct. 1968).

Salton, G., Search Strategy and the Optimization of Retrieval Effectiveness, in Mechanized Information Storage, Retrieval and Dissemination, Proc. F.I.D./I.F.I.P. Joint Conf., Rome, Italy, June 14-17, 1967, Ed. K. Samuelson, pp. 73-107 (North-Holland Pub. Co., Amsterdam, 1968).

Salton, G., The SMART Project - Status Report and Plans, in Information Storage and Retrieval, Rept. No. ISR-12, [G. Salton, Proj. Dir.] pp. I-1 to I-12 (Dept. of Computer Science, Cornell Univ., Ithaca, N.Y., June 1967).

Salton, G., The SMART System - Retrieval Results and Future Plans, in Information Storage and Retrieval, Rept. No. ISR-11, [G. Salton, Proj. Dir.] pp. I-1 to I-9 (Dept. of Computer Science, Cornell Univ., Ithaca, N.Y., June 1966).

Salton, G., The Use of Punctuation Patterns in Machine Translation, Mech. Trans. 5, No. 1, 16-24 (July 1958).

Salton, G. and M.E. Lesk, Computer Evaluation of Indexing and Text Processing, J. ACM 15, No. 1, 8-36 (Jan. 1968).

Salton, G. and M.E. Lesk, Information Analysis and Dictionary Construction, in Information Storage and Retrieval, Rept. No. ISR-11, [G. Salton, Proj. Dir.] pp. IV-1 to IV-71 (Dept. of Computer Science, Cornell Univ., Ithaca, N.Y., June 1966).

Salton, G., E.M. Keen and M. Lesk, Design Experiments in Automatic Information Retrieval, in The Growth of Knowledge, Ed. M. Kochen, pp. 336-351 (Wiley, New York, 1967).

Salton, G. and D.K. Williamson, A Comparison Between Manual and Automatic Indexing Methods, in Information Storage and Retrieval, Rept. No. ISR-14, [G. Salton, Proj. Dir.] pp. VI-1 to VI-44 (Dept. of Computer Science, Cornell Univ., Ithaca, N.Y., Oct. 1968).

Sams, B.H., On the Solution of an Information Retrieval Problem, AFIPS Proc. Spring Joint Computer Conf., Vol. 23, Detroit, Mich., May 1963, pp. 289-297 (Spartan Books, Baltimore, Md., 1963).

Samuelsdorff, P.O., The Participle in Modern Hebrew --- A Study in Automatic Ambiguity Resolution, in Computation in Linguistics: A Case Book, Ed. P.L. Garvin and B. Spolsky, pp. 252-283 (Indiana Univ. Press, Bloomington, 1966).

Samuelson, K., Ed., Mechanized Information Storage, Retrieval and Dissemination, Proc. F.I.D./I.F.I.P. Joint Conf., Rome, Italy, June 14-17, 1967, 729 p. (North-Holland Pub. Co., Amsterdam, 1968).

Saracevic, T., Quo Vadis Test and Evaluation, in Levels of Interaction Between Man and Information, Proc. Am. Doc. Inst. Annual Meeting, Vol. 4, New York, N.Y., Oct. 22-27, 1967, pp. 100-104 (Thompson Book Co., Washington, D.C., 1967).

Sasamori, K., M. Hashimoto, S. Ito and T. Yamazaki, Japanese Keyword Indexing Simulator (JAKIS) System III. Statistical Association Method, Paper III e 3, 33rd Conf. F.I.D. and Int. Cong. on Documentation, Tokyo, Sept. 12-22, 1967, 14 p. (preprint).

Satterthwait, A.C., Sentence-for-Sentence Translation: An Example, Mech. Trans. 8, 14-38 (Feb. 1965).

Savage, T.R., The Unevaluation of Automatic Indexing and Classification, (Abstract), in Statistical Association Methods For Mechanized Documentation, Symp. Proc., Washington, D.C., Mar. 17-19, 1964, NBS Misc. Pun. 269, Ed. M.E. Stevens et al, p. 211 (U.S. Govt. Print. Off., Washington, D.C., Dec. 15, 1965).

Savitt, D.A., H.H. Love, Jr. and R.E. Troop, ASP: A New Concept in Language and Machine Organization, AFIPS Proc. Spring Joint Computer Conf., Vol. 30, Atlantic City, N.J., April 18-20, 1967, pp. 87-102 (Thompson Books, Washington, D.C., 1967).

Schank, R.C. and L.G. Tesler, A Conceputal Parser for Natural Language, Proc. Int. Joint Conf. on Artificial Intelligence, Washington, D.C., May 7-9, 1969, Ed. D.E. Walker and L.M. Norton, pp. 569-578 (Int. Joint Conf. on Artificial Intelligence, 1969).

Schecter, G., Ed., Information Retrieval - A Critical View, Based on Third Annual Colloquium on Information Retrieval, Philadelphia, Pa., May 12-13, 1966, 282 p. (Thompson Book Co., Washington, D.C., 1967).

Scheffler, F.L. and R.B. Smith, Document Retrieval System Operations Including the Use of Microfiche and the Formulation of a Computer Aided Indexing Concept, Tech. Rept. No. AFML-TR-68-367, 50 p. (Air Force Materials Lab., Wright-Patterson AFB, Ohio, Feb. 1969).

Schiro, H., KWIC-Index, ein Maschinelles Dokumentations-Verfahren, (KWIC-Index, A Mechanical Procedure for Documentation), IBM Nachrichten 16, No. 177, 124-126 (Apr. 1966).

Schneider, K., Kie Herstellung von Stichwort-Registern (Construction of Key-Word Indexes), Nachrichten für Dokumentation 17, No. 5, 175-176 (Oct. 1966).

Schneider, K., UDC in Mechanized Indexing and Information Retrieval, in Mechanized Information Storage, Retrieval and Dissemination, Proc. F.I.D./I.F.I.P. Joint Conf., Rome, Italy, June 14-17, 1967, Ed. K. Samuelson, pp. 153-159 (North-Holland Pub. Co., Amsterdam, 1968).

Schnelle, H., Machine Translation of Languages --- A Critical Survey, Part 1, Sprachkunde und Informationsverarbeitung, No. 3, 41-61 (May 1964; Part 2, Ibid, No. 4, 58-63 (Nov. 1964) [German].

Schnelle, H., On the State of Research in Automatic Language Processing in German Speaking Areas, Sprachkunde und Informationsverarbeitung, No. 2, 48-61 (Nov. 1963) [German].

Schultz, C.K., Ed., H.P. Luhn: Pioneer of Information Science - Selected Works, 320 p. (Spartan Books, New York, 1968).

Schultz, C.K., An Imaginary Panel Discussion About Indexing, in Parameters of Information Science, Proc. Am. Doc. Inst. Annual Meeting, Vol. 1, Philadelphia, Pa., Oct. 5-8, 1964, pp. 437-452 (Spartan Books, Washington, D.C., 1964).

Schultz, C.K., W.L. Schultz and R.H. Orr, Comparative Indexing Terms Supplied by Biomedical Authors and by Document Titles, Am. Doc. 16, No. 4, 299-312 (Oct. 1965).

Schultz, L., Language and the Computer, in Automated Language Processing, Ed. H. Borko, pp. 11-31 (Wiley, New York, 1967).

Schwarcz, R. M., Steps Toward a Model of Linguistic Performance: A Preliminary Sketch, Research Memo. No. RM-5214-PR (The RAND Corp., Santa Monica, Calif., Jan. 1967).

Schwartz, J. T., Ed., Mathematical Aspects of Computer Science, Proc. Symposia in Applied Mathematics, Vol. 19, New York, N.Y., Apr. 5-7, 1966, 224 p. (American Mathematic Society, Providence, R.I., 1967).

Science Citation Index, 1966, Parts 9-14: Permuterm Subject Index, 6 vols. (Institute for Scientific Information, Philadelphia, Pa., 1967).

Sedano, J. M., Keyword-in-Context (KWIC) Indexing: Background, Statistical Evaluation, Pros and Cons, and Applications, M.S. Thesis, Pittsburgh Univ., Pa., 1964, 77 p.

Sedelow, S.Y. and W.A. Sedelow, Jr., Stylistic Analysis, in Automated Language Processing, Ed. H. Borko, pp. 181-213 (Wiley, New York, 1967).

See, R., Machine Aided Translation and Information Retrieval, in Electronic Handling of Information: Testing and Evaluation, Ed. A. Kent et al, pp. 89-108 (Thompson Book Co., Washington, D.C., 1967).

See, R., Mechanical Translation and Related Research, Science 144, 621-626 (May 1964).

Seidel, M., Threaded Term Association Files, in Statistical Association Methods For Mechanized Documentation, Symp. Proc., Washington, D.C., Mar. 17-19, 1964, NBS Misc. Pub. 269, Ed. M.E. Stevens et al, pp. 173-176 (U.S. Govt. Print. Off., Washington, D.C., Dec. 15, 1965).

Senechalle, D., Experiments with a New Classification Algorithm, Rept. No. LRC-64-WTM-6, 1 v. (Linguistics Research Center, Univ. of Texas, Austin, Dec. 1964).

Shannon, R.L., Experiment in Semiautomatic Indexing, USAEC: Appended to Research and Development Abstracts of the USAEC, RDA-3, pp. 1-45 (U.S. Atomic Energy Commission, Div. of Technical Information Extension, Oak Ridge, Tenn., July-Sept. 1962).

Shapiro, S.C. and G.H. Woodmansee, A Net Structure Based Relational Question Answerer: Description and Examples, Proc. Int. Joint Conf. on Artificial Intelligence, Washington, D.C., May 7-9, 1969, Ed. D.E. Walker and L.M. Norton, pp. 325-346 (Int. Joint Conf. on Artificial Intelligence, 1969).

Sharp, J.R., Content Analysis, Specification, and Control, in Annual Review of Information Science and Technology, Vol. 2, Ed. C.A. Cuadra, pp. 87-122 (Interscience Pub., New York, 1967).

Shastri, M.I., A Linguistic Approach to Relevance Judgment, Comparative Systems Lab. Tech. Rept. No. 12, 15 p. (Center for Documentation and Communication Research, Case Western Reserve Univ., Cleveland, O., July 1967).

Shaw, T.N., A Computer-Assisted Flexible Information Retrieval System, Aslib Proc. 20, No. 1, 34-39 (Jan. 1968).

Sherry, M.E., Syntactic Analysis in Automatic Translation, in Mathematical Linguistics and Automatic Translation, Rept. No. NSF-5, 1 v. (Harvard Univ., Cambridge, Mass., Aug. 1960).

Sieburg, J., Automatic Abstraction of Legal Information, Datamation 12, No. 11, 63-65 (Nov. 1966).

Silvano, A., Ed., American Handbook of Psychiatry, Vol. 3 (Basic Books, Inc., 1959).

Simmons, R. F., Answering English Questions by Computer: A Survey, Commun. ACM 8, 53-70 (Jan. 1965).

Simmons, R. F., Automated Language Processing, in Annual Review of Information Science and Technology, Vol. 1, Ed. C.A. Cuadra, pp. 137-169 (Interscience Pub., New York, 1966).

Simmons, R. F., Natural-Language Processing, Datamation 12, No. 6, 61-63, 65, 67, 69, 71-72 (June 1966).

Simmons, R. F., Natural-Language Processing by Computers - 1966, 31 p. (System Development Corp., Santa Monica, Calif., Dec. 1, 1965).

Simmons, R. F., Natural Language Processing and the Time-Shared Computer, in Toward a National Information System, Second Annual National Colloquium on Information Retrieval, Philadelphia, Pa., April 23-24, 1965, Ed. M. Rubinoff, pp. 217-227 (Spartan Books, Washington, D.C., 1965).

Simmons, R. F., Storage and Retrieval of Aspects of Meaning in Directed Graph Structures, Commun. ACM 9, 211-215 (Mar. 1966).

Simmons, R. F., J. F. Burger and R. E. Long, An Approach Toward Answering English Questions from Text, AFIPS Proc. Fall Joint Computer Conf., Vol. 29, San Francisco, Calif., Nov. 7-10, 1966, pp. 357-363 (Spartan Books, Washington, D.C., 1966).

Simmons, R. F., J. F. Burger and R. M. Schwarcz, A Computational Model of Verbal Understanding, AFIPS Proc. Fall Joint Computer Conf., Vol. 33, Pt. 1, San Francisco, Calif., Dec. 9-11, 1968, pp. 441-456 (Thompson Book Co., Washington, D.C., 1968).

Simonton, W., Ed., Information Retrieval Today, papers presented at the Institute conducted by the Library School of the Center for Continuation Study, Univ. of Minnesota, Minneapolis, Sept. 19-22, 1962, 176 p. (Univ. of Minnesota, Minneapolis, 1963).

Simonton, W. and C. Mason, Eds., Information Retrieval with special reference to the Biomedical Sciences, papers presented at the Second Institute on Information Retrieval, Univ. of Minnesota, Minneapolis, Nov. 10-13, 1965, 199 p. (Univ. of Minnesota, Minneapolis, 1966).

Skelly, S. J., Computerization of Canadian Statute Law, Law and Computer Tech. 1, No. 2, 10-14 (Feb. 1968).

Slagle, J. R., Experiments with a Deductive Question-Answering Program, Commun. ACM 8, 792-798 (Dec. 1965).

Slamecka, V., Ed., The Coming Age of Information Technology, Studies in Coordinate Indexing, Vol. 6, 166 p. (Documentation, Inc., Bethesda, Md., 1965).

Slamecka, V., Principles of Substantive Analysis of Information, Proc. 1965 Congress F.I.D. 31st Meeting and Congress, Vol. II, Washington, D.C., Oct. 7-16, 1965, pp. 229-234 (Spartan Books, Washington, D.C., 1966).

Slamecka, V. and P. Zunde, Automatic Subject Indexing from Contextual Condensations, in The Coming Age of Information Technology, Ed. V. Slamecka, pp. 114-121 (Documentation, Inc., Bethesda, Md., 1965).

Sneath, P.H.A., A Comparison of Different Clustering Methods as Applied to Randomly-Spaced Points, Classification Soc. Bull. 1, No. 2, 2-18 (1966).

Soergel, D., Mathematical Analysis of Documentation Systems, An Attempt to a Theory of Classification and Search Request Formulation, Inf. Stor. & Retr. 3, No. 3, 129-173 (July 1967).

Solomonoff, R.J., A Progress Report on Machines to Learn to Translate Languages and Retrieve Information, Rept. No. ZTB-134, 17 p. (Zator Co., Cambridge, Mass., Oct. 1959).

Spärck-Jones, K., Automatic Term Classification and Information Retrieval, Proc. IFIP Congress 68, Edinburgh, Scotland, Aug. 5-10, 1968, Booklet G, pp. G 5 - G 9 (North-Holland Pub. Co., Amsterdam, 1968).

Spärck-Jones, K., Experiments in Semantic Classification, Mech. Trans. 8, No. 3-4, 97-112 (June-Oct. 1965).

Spärck-Jones, K., Synonymy and Semantic Classification, Rept. No. ML-170, 258 p. (Cambridge Language Research Unit, Cambridge, England, June 1964).

Spärck-Jones, K. and D. Jackson, Current Approaches to Classification and Clump-Finding at the Cambridge Language Research Unit, The Computer J. 10, 29-37 (May 1967).

Spärck-Jones, K. and D. Jackson, Some Experiments in the Use of Automatically-Obtained Term Clusters for Retrieval, in Mechanized Information Storage, Retrieval and Dissemination, Proc. F.I.D./I.F.I.P. Joint Conf., Rome, Italy, June 14-17, 1967, Ed. K. Samuelson, pp. 203-212 (North-Holland Pub. Co., Amsterdam, 1968).

Spärck-Jones, K. and D.M. Jackson, The Use of the Theory of Clumps for Information Retrieval: Report on the O.S.T.I.-Supported Project at the Cambridge Language Research Unit, Rept. No. ML-200, 1 v. (Cambridge Language Research Unit, Cambridge, England, 1966).

Spärck-Jones, K. and R.M. Needham, Automatic Term Classifications and Retrieval, Inf. Stor. & Retr. 4, No. 2, 91-100 (June 1968).

Spencer, C.C., Subject Searching with Science Citation Index: Preparation of a Drug Bibliography Using Chemical Abstracts, Index Medicus, and Science Citation Index 1961 and 1964, Am. Doc. 18, No. 2, 87-95 (April 1967).

Spiegel, J. and E. Bennett, A Modified Statistical Association Procedure for Automatic Document Content Analysis and Retrieval, in Statistical Association Methods For Mechanized Documentation, Symp. Proc., Washington, D.C., Mar. 17-19, 1964, NBS Misc. Pub. 269, Ed. M.E. Stevens et al, pp. 47-60 (U.S. Govt. Print. Off., Washington, D.C., Dec. 15, 1965).

Stevens, M.E., Automatic Analysis, in Encyclopedia of Library and Information Science, Vol. 2, Ed. A. Kent and H. Lancour, pp. 144-184 (Marcel Dekker, Inc., New York, 1969).

Stevens, M.E., Nonnumeric Data Processing in Europe: A Field Trip Report, August-October 1966, NBS Tech. Note 462, 62 p. (U.S. Govt. Print. Off., Washington, D.C., Nov. 1968).

Stevens, M.E., Problems and Prospects in Mechanized Indexing, invited paper, Symp. on Mechanized Abstracting and Indexing - Papers and Discussion, Moscow, Sept. 28 - Oct. 1, 1966, pp. 4-24, Unesco Doc. No. SC/WS/172, issued Paris, Jan. 12, 1968 (Distribution Limited).

Stevens, M.E. and G.H. Urban, Automatic Indexing Using Cited Titles, in Statistical Association Methods For Mechanized Documentation, Symp. Proc., Washington, D.C., Mar. 17-19, 1964, NBS Misc. Pub. 269, Ed. M.E. Stevens et al, pp. 213-215 (U.S. Govt. Print. Off., Washington, D.C., Dec. 15, 1965).

Stevens, M.E., V.E. Giuliano and L.B. Heilprin, Eds., Statistical Association Methods For Mechanized Documentation, Symp. Proc., Washington, D.C., Mar. 17-19, 1964, NBS Misc. Pub. 269, 261 p. (U.S. Govt. Print. Off., Washington, D.C., Dec. 15, 1965).

Stickel, G., Automatische Textzerlegung und Registerherstellung, Rept. No. PI-11, Deutsches Reckenzentrum, Darmstadt, Federal Republic of Germany, Dec. 1964, 16 p.

Stiles, H.E., Automatic Indexing and the Association Factor, in Information Systems Compatibility, based on papers given at the 6th Institute on Information Storage and Retrieval, American Univ., Washington, D.C., 1964, Ed. S.M. Newman, pp. 135-142 (Spartan Books, Washington, D.C., 1965).

Stitelman, J., International Cooperation in Automated Patent Searching, Law and Computer Tech. 1, No. 2, 15-19 (Feb. 1968).

Stogniy, A.A. and V.N. Afanassiev, Some Design Problems for Automatic Fact Information Retrieval and Storage Systems, in Mechanized Information Storage, Retrieval and Dissemination, Proc. F.I.D./I.F.I.P. Joint Conf., Rome, Italy, June 14-17, 1967, Ed. K. Samuelson, pp. 289-299 (North-Holland Pub. Co., Amsterdam, 1968).

Stone, D.C. and M. Rubinoff, Statistical Generation of a Technical Vocabulary, Am. Doc. 19, No. 4, 411-412 (Oct. 1968).

Stone, P.J., Transformation and Organization of Information Content: Contribution of Psychology, Proc. 1965 Congress F.I.D. 31st Meeting and Congress, Vol. II, Washington, D.C., Oct. 7-16, 1965, pp. 83-86 (Spartan Books, Washington, D.C., 1966).

Stone, P.J., D.C. Dunphy, M.S. Smith and D.M. Ogilvie, The General Inquirer, A Computer Approach to Content Analysis, 651 p. (M.I.T. Press, Cambridge, Mass., 1966).

Stone, P.J. and E.B. Hunt, A Computer Approach to Content Analysis: Studies Using the General Inquirer System, AFIPS Proc. Spring Joint Computer Conf., Vol. 23, Detroit, Mich., May 1963, pp. 241-256 (Spartan Books, Baltimore, Md., 1963).

Swanson, D.R., On Indexing Depth and Retrieval Effectiveness, in Second Cong. on the Information System Sciences, Hot Springs, Va., Nov. 1964, Ed. J. Spiegel and D.E. Walker, pp. 311-319 (Spartan Books, Washington, D.C., 1965).

Switzer, P., Vector Images in Document Retrieval, in Statistical Association Methods For Mechanized Documentation, Symp. Proc., Washington, D.C., Mar. 17-19, 1964, NBS Misc. Pub. 269, Ed. M.E. Stevens et al, pp. 163-171 (U.S. Govt. Print. Off., Washington, D.C., Dec. 15, 1965).

Szanser, A.J., Error-Correcting Methods in Natural Language Processing, Proc. IFIP Congress 68, Edinburgh, Scotland, Aug. 5-10, 1968, Booklet H, pp. H 15 - H 19 (North-Holland Pub. Co., Amsterdam, 1968).

Tabidze, G.S., Realization of Machine Algorithm for the Detection in Text of Object Names, unedited rough draft translation of Nauchno Tekhnicheskaya Informatsiia, No. 8, 47-50 (1964).

Tabidze, G.S., Realization of Machine Algorithm for the Detection in Text of Object Names, Rept. No. FTD-TT-65-1083, 18 p. (Foreign Technology Div., Wright-Patterson AFB, O., Jan. 3, 1966).

Tarnawsky, G.O., Tagging Techniques for Incorporating Microglossaries in an Automatic Dictionary, IBM J. Res. & Dev. 7, No. 4, 337-339 (Oct. 1963).

Taube, M., Extensive Relations as the Necessary Condition for the Significance of Thesauri for Mechanized Indexing, J. Chem. Doc. 3, No. 3, 177-180 (July 1963).

Taulbee, O.E., Content Analysis, Specification, and Control, in Annual Review of Information Science and Technology, Vol. 3, Ed. C.A. Cuadra, pp. 105-136 (Encyclopedia Britannica, Inc., Chicago, Ill., 1968).

Taulbee, O.E. and J.T. Welch, Jr., A New Classification Theory Leading to Automatic Pattern Recognition, Proc. 21st National Conf., ACM, Los Angeles, Calif., Aug. 30 - Sept. 1, 1966, pp. 63-67 (Thompson Book Co., Washington, D.C., 1966).

Tell, B., ABACUS [in Swedish], Biblioteksbladet 53, No. 1-2, 21-30 (1968).

Terzi, P., An Hypothetical Mechanism of the Origin of Ideas as an Instrument for the Automatic Analysis of Language [Un Ipotetico Meccanismo delle Origini delle Idee Quale Strumento per l'Analisi Automatica del Linguaggio], Istituto Lombardo, Accademia di Scienze e Lettere, Milan, Feb. 17, 1966.

Terzi, P., The Value of Mathematics in Resolving the Problem of Machine Translation, Automaz. Automat. (Milan) 10, 13-20 (Mar.-Apr. 1966). (Italian version appended, 8 p.).

Tharp, A.L. and G.K. Krulee, Using Relational Operators to Structure Long-Term Memory, Proc. Int. Joint Conf. on Artificial Intelligence, Washington, D.C., May 7-9, 1969, Ed. D.E. Walker and L.M. Norton, pp. 579-586 (Int. Joint Conf. on Artificial Intelligence, 1969).

Thomas, C.B., An Automatic Phrase Structure Analysis of a Spanish Text, Rept. No. LRC-65-WD-2, 134 p. (Linguistics Research Center, Univ. of Texas, Austin, Sept. 1965).

Thompson, F.B., English for the Computer, AFIPS Proc. Fall Joint Computer Conf., Vol. 29, San Francisco, Calif., Nov. 7-10, 1966, pp. 349-356 (Spartan Books, Washington, D.C., 1966).

Tinker, J.F., Imprecision in Indexing, Part II, Am. Doc. 19, No. 3, 322-330 (July 1968).

Tinker, J.F., Imprecision in Meaning Measured by Inconsistency of Indexing, Am. Doc. 17, No. 2, 96-102 (April 1966).

Tomlinson, H., Classification of Information Topics by Clustering Interest Profiles, Rept. No. PRL-TR-65-19, 18 p. (Aerospace Medical Div., Lackland AFB, Texas, Nov. 1965).

Tompkins, M.L. and J.W. Tukey, Permuted (Circularly-Shifted) Indexes to Abbreviations: A Mechanically Prepared Aid to Serial Identification, in Progress in Information Science and Technology, Proc. Am. Doc. Inst., Annual Meeting, Vol. 3, Santa Monica, Calif., Oct. 3-7, 1966, pp. 347-355 (Adrianne Press, 1966).

Tosh, W., Syntactic Translation, 162 p. (Mouton, The Hague, 1965).

Tou, J.T., Ed., Computer and Information Sciences-II, Proc. 2nd Symp. on Computer and Information Sciences, Columbus, O., Aug. 22-24, 1966, 368 p. (Academic Press, New York, 1967).

Tou, J.T. and R.H. Wilcox, Eds., Computer and Information Sciences, collected papers on learning, and adaptation and control in information systems, 544 p. (Spartan Books, Washington, D.C., 1964).

Travis, L.E., Analytic Information Retrieval, in Natural Language and the Computer, Ed. P.L. Garvin, pp. 310-353 (McGraw-Hill, New York, 1963).

Treu, S., The Browser's Retrieval Game, Am. Doc. 19, No. 4, 404-410 (Oct. 1968).

Uhlmann, W., Computerized Concert Index of Swedish Archives of Music History, An Example for the Retrieval of Historical Data, FOA Index Rept. 0616-10, 1 v. (Swedish Research Institute for National Defense, Stockholm, 1967).

Ullman, J.R., Algebraic Inference of Pattern Similarity, The Computer J. 10, No. 3, 256-264 (Nov. 1967).

United Nations Educational, Scientific and Cultural Organization, Symp. on Mechanized Abstracting and Indexing - Final Report, UNESCO/NS/209, Paris, April 28, 1967, 6 p.

United Nations Educational, Scientific and Cultural Organization, Symp. on Mechanized Abstracting and Indexing - Papers and Discussion, Moscow, Sept. 28 - Oct. 1, 1966, 51 p., Unesco Doc. No. SC/WS/172, issued Paris, Jan. 12, 1968, (Distribution Limited).

Vann, J.O. and C.L. Bernier, Letter to the Editor, Am. Doc. 19, No. 1, 105-106 (Jan. 1968).

Vasarhelyi, P.E., Project Transinform: A Computer Based Machine Indexing System with Simple Machine Translation, and the International Version of it, Using Esperanto as a Common Intermediate Language, Proc. 33rd Conf. FID and Int. Congress on Documentation, Tokyo, Sept. 1967. Abstracts [Tokyo] 1967, p. 38.

Vaswani, P.K.T., Mechanized Storage and Retrieval of Information, Rev. Int. Doc. 32, 19-22 (1965).

Vaswani, P.K.T., A Technique for Cluster Emphasis and Its Application to Automatic Indexing, Proc. IFIP Congress 68, Edinburgh, Scotland, Aug. 4-10, 1968, Booklet G, pp. G 1 - G 4 (North-Holland Pub. Co., Amsterdam, 1968).

Veillion, G. and J. Veyrunes, Etude de la Réalisation Pratique d'une Grammaire ≪Context-Free≫ et de l'Algorithme Associé, La Traduction Automatique 5, No. 3, 69-79 (Sept. 1964).

Vejsová, A., Preparation of Automation in Universally Technical Collections: Its Experimental and Analytical Phase, in Mechanized Information Storage, Retrieval and Dissemination, Proc. F.I.D./I.F.I.P. Joint Conf., Rome, Italy, June 14-17, 1967, Ed. K. Samuelson, pp. 485-497 (North-Holland Pub. Co., Amsterdam, 1968).

Venezky, R.L., Computer-Aided Humanities Research at the University of Wisconsin, Computers and the Humanities 3, No. 3, 129-138 (Jan. 1969).

Vickery, B.C., On Retrieval System Theory, 2nd ed., 191 p. (Butterworths, Washington, D.C., 1965).

Vleduts, G.E., V.V. Nalimov and N.I. Styazhkin, Scientific and Technical Information as one of the Problems of Cybernetics, Soviet Physics Uspekhi 2 (69), No. 5, 637-665 (Sept.-Oct. 1959).

VonBriesen, R., Status of Legal Use of Computers, Law and Computer Tech. 1, No. 4, 9-18 (April 1968).

VonGlasersfeld, E., "Multistore", A Procedure for Correlational Analysis, IDAMI Language Research Section, Rept. No. ILRS-T10-650120, 87 p. (Istituto Documentziaone della Associazione Meccanica Italiana, Milan, Italy, Jan. 1965).

VonGlasersfeld, E., Problems of Machine Translation, Sprachkunde und Informationsverarbeitung, No. 2, 33-47 (Nov. 1963). [German].

Von Glasersfeld, E., A Project for Automatic Sentence Analysis, Rept. No. ILRS-T3, 640331, 102 p. (Instiuto Documentazione dell' Associazion: Meccanica Italiana, Milan, Sept. 1966).

Von Glasersfeld, E., J. Burns, P.P. Pisani, B. Notarmarco and B. Dutton, Automatic English Sentence Analysis, 111 p. (IDAMI Language Research Center, Milan, Sept. 1966).

Wagner, S.W., Automatische Stichwortanalyse nach dem Rangkriterienverfahren, Doctoral Dissertation, Fakultät für Elektrontechnik dr Technischen Hochschule Karlsruhe, Karlsruhe, Federal Republic of Germany, 1966, 116 p.

Walker, D.E., SAFARI, An On-Line Text-Processing System, in Levels of Interaction Between Man and Information, Proc. Am. Doc. Inst. Annual Meeting, Vol. 4, New York, N.Y., Oct. 22-27, 1967, pp. 144-147 (Thompson Book Co., Washington, D.C., 1967).

Walker, D.E., Recent Developments in the MITRE Syntactic Analysis Procedure, Rept. No. MTP-11, 47 p. (MITRE Corp., Bedford, Mass., Sept. 1966).

Walker, D.E. and L.M. Norton, Eds., Proc. International Joint Conf. on Artificial Intelligence, Washington, D.C., May 7-9, 1969, 715 p. (Int. Joint Conf. on Artificial Intelligence, 1969).

Wallace, E.M., Rank Order Patterns of Common Words as Discriminators of Subject Content in Scientific and Technical Prose, in Statistical Association Methods For Mechanized Documentation, Symp. Proc., Washington, D.C., Mar. 17-19, 1964, NBS Misc. Pub. 269, Ed. M.E. Stevens et al, pp. 225-229 (U.S. Govt. Print. Off., Washington, D.C., Dec. 15, 1965).

Wang, T.L., An Information System with the Ability to Extract Intelligence from Data, Commun. ACM 5, 16-18 (Jan. 1962).

Weber, R.W., Associate Processing of Fragmentary Information, IEEE Trans. EWS-8, 71-80 (Dec. 1965).

Weizenbaum, J., Contextual Understanding by Computers, Commun. ACM 10, No. 8, 474-480 (Aug. 1967).

Weizenbaum, J., ELIZA - A Computer Program for the Study of Natural Language Communication Between Man and Machine, Commun. ACM 9, 36-45 (Jan. 1966).

Welt, I.D., Indexes and Index Mechanization in Biomedicine, J. Chem. Doc. 3, 169-174 (1963).

Williams, J.H. Jr., BROWSER - An Automatic Indexing On-Line Text Retrieval System, Annual Progress Report, Contract NONR 4456(00), 28 p. (IBM Corp., Federal Systems Div., Gaithersburg, Md., Sept. 1969).

Williams, J.H. Jr., Computer Classification of Documents, in Mechanized Information Storage, Retrieval and Dissemination, Proc. F.I.D./I.F.I.P. Joint Conf., Rome, Italy, June 14-17, 1967, Ed. K. Samuelson, pp. 235-246 (North-Holland Pub. Co., Amsterdam, 1968).

Williams, J.H. Jr., Discriminant Analysis for Content Classification, 272 p. (IBM Corp., Bethesda, Md., Dec. 1965).

Williams, J.H. Jr., Results of Classifying Documents with Multiple Discriminant Functions, in Statistical Association Methods For Mechanized Documentation, Symp. Proc., Washington, D.C., Mar. 17-19, 1964, NBS Misc. Pub. 269, Ed. M.E. Stevens et al, pp. 217-224 (U.S. Govt. Print. Off., Washington, D.C., Dec. 15, 1965).

Williams, J.H. Jr. and M.P. Perriens, Automatic Full Text Indexing and Searching System, in IBM Proc. Inf. Systems Symp., Washington, D.C., Sept. 4-6, 1968, pp. 335-350 (IBM Corp., 1968).

Winters, W.K., A Modified Method of Latent Class Analysis for File Organization in Information Retrieval, J. ACM 12, No. 3, 356-363 (July 1965).

Woods, W.A., Procedural Semantics for a Question-Answering Machine, AFIPS Proc. Fall Joint Computer Conf., Vol. 33, Pt. 1, San Francisco, Calif., Dec. 9-11, 1968, pp. 457-471 (Thompson Book Co., Washington, D.C., 1968).

Wyllys, R.E., Extracting and Abstracting by Computer, in Automated Language Processing, Ed. H. Borko, pp. 127-179 (Wiley, New York, 1967).

Yakushin, B.V., Problems of Algorithmic Composition of Subject Indexes (Brief Survey of Foreign Literature), LT-65-102, Mar. 15, 1966, 18 p. Transl. of Nauchno-Tekhnicheskaya Informatsiya (USSR) n. 5, 22-25 (1965).

Yamada, S., Mechanical Syntactic Analysis, Information Processing in Japan 5, 24-26 (1965).

Yeats, J.C.R., The Statement Index: A Subject Index Constructed from the Syntax of Titles, in Looking Forward in Documentation, Section 2, 43 p. (Aslib, London, 1965).

Yngve, V.H., A Framework for Syntactic Translation, in Readings in Automatic Language Processing, Ed. D.G. Hays, pp. 189-198 (American Elsevier Pub., New York, 1966).

Zint, I., On the Present State of Automated Language Processing, Linguistik und Informationsverarbeitung, No. 12, 36-55 (Dec. 1967).

Zunde, P., Automatic Indexing from Machine Readable Abstracts of Scientific Documents, Rept. No. AFOSR 65-1425, 213 p. (Documentation, Inc., Bethesda, Md., Sept. 1965).

Zunde, P. and M.E. Dexter, Indexing Consistency and Quality, Am. Doc. 20, No. 3, 259-267 (July 1969).

Zunde, P. and V. Slamecka, Distribution of Indexing Terms from Maximum Efficiency of Information Transmission, Am. Doc. 18, No. 2, 104-108 (April 1967).

Zunde, P., F.T. Armstrong and T.T. Stretch, Evaluating and Improving Internal Indexes, in Levels of Interaction Between Man and Information, Proc. Am. Doc. Inst. Annual Meeting, Vol. 4, New York, N.Y., Oct. 22-27, 1967, pp. 86-89 (Thompson Book Co., Washington, D.C., 1967).

Zwicky, A.M., J. Friedman, B.C. Hall and D.E. Walker, The MITRE Syntactic Analysis Procedure for Transformational Grammars, AFIPS Proc. Fall Joint Computer Conf., Vol. 27, Pt. 1, Las Vegas, Nev., Nov. 30 - Dec. 1, 1965, pp. 317-326 (Spartan Books, Washington, D.C., 1965).

☆ U. S. GOVERNMENT PRINTING OFFICE : 1970 O - 381-006

NBS TECHNICAL PUBLICATIONS

PERIODICALS

JOURNAL OF RESEARCH reports National Bureau of Standards research and development in physics, mathematics, chemistry, and engineering. Comprehensive scientific papers give complete details of the work, including laboratory data, experimental procedures, and theoretical and mathematical analyses. Illustrated with photographs, drawings, and charts.

Published in three sections, available separately:

● Physics and Chemistry

Papers of interest primarily to scientists working in these fields. This section covers a broad range of physical and chemical research, with major emphasis on standards of physical measurement, fundamental constants, and properties of matter. Issued six times a year. Annual subscription: Domestic, $9.50; foreign, $11.75*.

● Mathematical Sciences

Studies and compilations designed mainly for the mathematician and theoretical physicist. Topics in mathematical statistics, theory of experiment design, numerical analysis, theoretical physics and chemistry, logical design and programming of computers and computer systems. Short numerical tables. Issued quarterly. Annual subscription: Domestic, $5.00; foreign, $6.25*.

● Engineering and Instrumentation

Reporting results of interest chiefly to the engineer and the applied scientist. This section includes many of the new developments in instrumentation resulting from the Bureau's work in physical measurement, data processing, and development of test methods. It will also cover some of the work in acoustics, applied mechanics, building research, and cryogenic engineering. Issued quarterly. Annual subscription: Domestic, $5.00; foreign, $6.25*.

TECHNICAL NEWS BULLETIN

The best single source of information concerning the Bureau's research, developmental, cooperative and publication activities, this monthly publication is designed for the industry-oriented individual whose daily work involves intimate contact with science and technology—*for engineers, chemists, physicists, research managers, product-development managers, and company executives.* Annual subscription: Domestic, $3.00; foreign, $4.00*.

*** Difference in price is due to extra cost of foreign mailing.**

NONPERIODICALS

Applied Mathematics Series. Mathematical tables, manuals, and studies.

Building Science Series. Research results, test methods, and performance criteria of building materials, components, systems, and structures.

Handbooks. Recommended codes of engineering and industrial practice (including safety codes) developed in cooperation with interested industries, professional organizations, and regulatory bodies.

Special Publications. Proceedings of NBS conferences, bibliographies, annual reports, wall charts, pamphlets, etc.

Monographs. Major contributions to the technical literature on various subjects related to the Bureau's scientific and technical activities.

National Standard Reference Data Series. NSRDS provides quantitive data on the physical and chemical properties of materials, compiled from the world's literature and critically evaluated.

Product Standards. Provide requirements for sizes, types, quality and methods for testing various industrial products. These standards are developed cooperatively with interested Government and industry groups and provide the basis for common understanding of product characteristics for both buyers and sellers. Their use is voluntary.

Technical Notes. This series consists of communications and reports (covering both other agency and NBS-sponsored work) of limited or transitory interest.

Federal Information Processing Standards Publications. This series is the official publication within the Federal Government for information on standards adopted and promulgated under the Public Law 89–306, and Bureau of the Budget Circular A–86 entitled, Standardization of Data Elements and Codes in Data Systems.

CLEARINGHOUSE

The Clearinghouse for Federal Scientific and Technical Information, operated by NBS, supplies unclassified information related to Government-generated science and technology in defense, space, atomic energy, and other national programs. For further information on Clearinghouse services, write:

Clearinghouse
U.S. Department of Commerce
Springfield, Virginia 22151

Order NBS publications from:

Superintendent of Documents
Government Printing Office
Washington, D.C. 20402